Survivors crossed the K-Pg extinction line and radiated in the Quaternary: Phylogenetic genomics of the Cimbicidae

Gengyun Niu[1], Yalan Cheng[1], Yuchen Yan[2], Zemin Sun[1], Siying Wan[1] & Meicai Wei[1,*]

[1] College of Life Sciences, Jiangxi Normal University, Nanchang, Jiangxi, China

[2] College of Life Sciences, Central South University of Forestry and Technology, Changsha, China

*Corresponding Author:

Meicai Wei

College of Life Sciences, Jiangxi Normal University, 99 Ziyang Road, Nanchang, Jiangxi 330022, China

Email address: weim@jxnu.edu.cn

chinaXiv:202303.00021v1

Author contributions

Conceptualization, M.W.; methodology, M.W.; validation, Y.C.; formal analysis, Y.C., and G.N.; investigation, Y.Y., and G.N.; resources, M.W.; data curation, G.N.; writing-original draft, G.N., and M.W.; writing-review and editing, G.N.; visualization, S.W., and Z.S.; supervision, M.W.; project administration, M.W.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

chinaXiv:202303.00021v1

Abstract

The Cimbicidae are the physically largest members of the Hymenoptera. They are herbivorous sawflies with clubbed antennae. The previous classification maintained that this family contains four subfamilies. Two of them, the Cimbicinae and Abiinae, are richly diverse and distributed across the Holarctic. The Corynidinae are confined to the Palaearctic realm and primarily diversified around the Mediterranean and southwestern Asia, while the most morphologically primitive Pachylostictinae has a restricted distribution in South America. However, the connotation of these subfamilies and phylogenetic relationships among genera is still confusing, which limits the study of their evolutionary history and hypotheses of their particular origin. Here, we used the nuclear single-copy genes and mitochondrial genomes to trace the evolutionary history of the Cimbicidae and combine an extensive molecular dataset with phylogenetically and stratigraphically constrained fossil calibrations to deduce an evolutionary timescale for the Cimbicidae. We reveal that the Cimbicidae survived the Cretaceous-Palaeogene (K-Pg) extinction. After then, the lineages differentiation and the diversification of the extant genera gradually initiated in the earlier half of the Paleogene. However, the rapid diversification of the Cimbicidae was almost completed in the later half of Paleogene and the earlier half of Neogene (40–10 Myr.). This fast and almost simultaneous genus-level diversity of Cimbicidae underscores the significance of the boundary of geological historical sequence in shaping the taxonomic hierarchy of insects.

Keywords: [Single-copy genes] [phylogenomics] [mt-genome] [Cimbicidae] [K-Pg]

chinaXiv:202303.00021v1

1. Introduction

The Cimbicidae is a relatively small family of herbivores; yet it contains the largest specimens of true sawflies (Vilhelmsen et al., 2021); their size could offer them unique strategies for survival. The larvae are phytophagous and favor a variety of dicotyledon angiosperms (Goulet, 1992). The clubbed antenna is the most practical distinguishing feature of the Cimbicidae, although it occurs several times in the Tenthredinoidea, such as in some genera and species of the Tenthredinidae and Pergidae. In addition to the clubbed antenna with a much elongated third antennomere, the following could also be apomorphic characters for the family. These include abdomens with distinct lateral carina and at least some laterotergite that are distinctly separated from the main tergite by a suture. The first abdominal tergum fuses laterally with metapleuron and also fuses on the meson. The first abscissa of vein Rs absent hence the cells 1R1 and 1Rs merged, 1R+1Rs not shorter than 2Rs; the veins C and Sc+R very close and the free Sc1 absent; the vein cu-a close to the base of 1M; the vein R+M about as long as vein 1M; and the antenna of larvae consist of two antennomeres (Yan et al., unpublished).

Although the Cimbicidae only contain approximately 213 described extant species, the morphological diversity of differentiation of the 21 genera is clearly observable and easily recognized (Yang, et al., 2022; Yang, et al., 2021). The Cimbicinae and Abiinae are the most diversified with 15 genera and 191 species. This family primarily occurs in the Holarctic region, extending south to north of the Oriental region, and only the Pachylostictinae with five genera and nine species occurs in South America. The subfamily classification proposed by Benson (1938) is remarkably close to the current understanding of the Cimbicidae. Abe and Smith (1991)followed Benson's system and divided the subfamily Cimbicinae into two tribes: Cimbicini and Trichiosomini. The fossil taxa were also arranged into the new system. However, without any explanation, their treatment of several fossil genera, such as

Trichiosomites and *Phenacoperga*, has not been accepted by subsequent studies (e. g. Ren et al., 2020; Vilhelmsen, 2019). Following the system of Abe and Smith (1991), Wei et al. (2012)compiled a key to the genera and tribes of the family based on the Asian species. Recently, Vilhelmsen (2019) scored 144 morphological characters from the adult anatomy for 95 Cimbicids and 26 outgroup taxa and confirmed the classification scheme proposed by Benson (1938). However, the monophyly of Pachylostictinae had not been retrieved in all analyses, but it is inferred from its geographical distribution and morphological isolation. In summary, the classifications described above are not accepted by all experts. Although the composition of the Cimbicidae is clear, the hierarchical structuring into subfamilies remains a contentious issue.

With moderate species diversity, the Cimbicidae cover several diversified models and have typical patterns of distribution. These characteristics determine that this family should be a fascinating model to address fundamental phenotypic, molecular biogenetic, and biogeographical evolution and explore the patterns of insect evolution. However, the evolutionary history of Cimbicidae has rarely been studied. Early molecular divergence dating for all of the Hymenoptera (O'Reilly et al., 2015; Ronquist et al., 2012) shows that the stem and crown age of the Cimbicidae is 160~200 Ma and 135~110 Ma, respectively (Table S1) and suggests that the major clades of Cimbicidae diversified during the Upper Cretaceous period (Isaka et al., 2015; Niu et al., 2021; Nyman et al., 2019). The only analyses with sufficient samples to achieve subfamily-level resolution used *Cenocimbex menatensis* to limit the lower boundary of the crown Cimbicid and consider the divergence of Cimbicinae to have occurred in 35 Ma, and that of the Abiinae to have occurred in 20 Ma (Nyman et al., 2019). In contrast, the fossils are much younger. The oldest definitive fossil of the Cimbicidae, *Cenocimbex menatensis*, was reported from Menat, France, and dated to Selandian in the Paleocene period. Younger fossils are diverse and were widely distributed from Ypresian of

chinaXiv:202303.00021v1

Okanagan Highland, Canada, and Burdigalian of Shanwang, China, during the Miocene (Table S2). Although from different stratigraphic sources, the fossils have a highly homogeneous morphology, which makes most records useless for calibrating the phylogenetic node. In particular, the parallelism of veins in the subfamilies Abiinae and Corynidinae are only found in Shanwang, which is not enough to calibrate the origin of related groups.

In this study, we produce a phylogeny of the Cimbicidae to generate a hypothesis of the natural relationships in the family through phylogenetic analyses using the mitochondrial genome (mtG) and nuclear single-copy orthologs (SCOs) and by increasing sampling from East Asia, which had not been previously studied. We conducted fossil-derived calibrations combined with relaxed clock analyses to understand the temporal framework of the evolution. In this process, we performed different analyses with alternative positions of the oldest fossil. By exploring the causal events or processes in the underlying phylogenetic diversity, the possible scenarios of evolutionary history are inferred.

chinaXiv:202303.00021v1

2. Materials and methods

2.1 Taxon sampling

The analyses included 25 extant Cimbicids across three of four subfamilies and 15 of 21 genera (Table S3). *Asitrichiosoma anthracinum* (KT921411) and *Corynis lateralis* (KY063728) were sequenced by the Sanger method (by author communication), and thus, were excluded from the nuclear analyses.

In advance, the mtG phylogenetic tree was constructed using all the Cimbicids and 40 other Symphytan to obtain the framework and rooting strategy. Subsequent phylogenetic reconstruction removed all the outgroups to avoid errors caused by interfamily heterogeneity.

Samples were identified by Wei Meica. Voucher specimens have been deposited at the

Asia Sawfly Museum, Nanchang (ASMN).

2.2 Sequencing, annotation, and analyses

We performed whole-genome sequencing (WGS) using next-generation sequencing (NGS). The genomic DNA was pooled and sequenced with a high-throughput Illumina HiSeq 4000 platform (Illumina, Inc., San Diego, CA, USA) using 150 bp paired-end libraries. The mtGs were reconstructed using a combination of *de novo* and reference guide assemblies using MitoZ (Meng, et al., 2019) and Geneious Prime 2019.2.1 (https://www.geneious.com) (Kearse et al., 2012; Kumar et al., 2016), respectively. The genes for RNA of each mitogenome were identified based on their putative secondary structures using the MITOS web server (http://mitos.bioinf.uni-leipzig.de/index.py) (Bernt et al., 2013) with the invertebrate mitochondrial genetic code. The initiation and termination of protein-coding genes (PCGs) were determined by comparison with other symphytans species. The basic nucleotide composition and relative synonymous codon usage (RSCU) of the PCGs were calculated using the MEGA application v. 7.0 (Kumar et al., 2016). Strand asymmetry was computed using the formulae for the strand that encode the majority of PCGs: AT-skew = (A - T) / (A + T) and GC-skew = (G - C) / (G + C) (Perna et al., 1995).

The low-coverage genomes (LCG) were generated through rapid genome assemblies using the pipeline PLWS (http://github.com/xtmtd/PLWS, accessed on October 30, 2019), as described by Zhang et al. (2019). BBTools v. 37.93 (Bushnell, 2020), Lighter v. 1.1.1 (Song, et al., 2014), Minia v. 3.00-alpha1 (Chikhi et al., 2012), Redundans v. 0.13c (Pryszcz et al., 2016), BESST v. 2.2.8 (Sahlin et al., 2014), GapCloser v. 1.12 (Luo, R. et al., 2012) and GenomeScope v. 1.0.0 (Vurture et al., 2017) were used in the pipeline to assembly the genome. The reassembly using SPADES (Bankevich et al., 2012) was conducted when the Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness score was < 80%.

We used BUSCO v. 3.0.2 (Simão et al., 2015) with hymenoptera_odb10 (n= 5991)

chinaXiv:202303.00021v1

(Manni et al., 2021) to retrieve SCOs. The nucleotide sequences of all the orthologs were translated to amino acid (aa) sequences. All the orthologs were aligned by MAFFT v. 7.182 (Katoh et al., 2008) based on their aa sequence using L-INS-i. Next, PAL2NAL (Suyama et al., 2006) was used to translate the aa sequence alignments to codon sequence alignments, and trimAl (Capella-Gutiérrez et al., 2009) was used with "automated1" to trim the aa sequence alignments. The trimmed segments of the aa sequence alignments were deleted from their corresponding codon sequence alignments using custom Perl scripts. BaCoCa (Kück et al., 2014) was used to detect the compositional heterogeneity and bias (RCFV value), and then the aa with RCFV values < 0.1 were selected to reconstruct the phylogenetic tree.

2.3 Secondary structure prediction

The secondary structures of 12S rRNA and 16S rRNA were partitioned into four and six areas, respectively. The secondary structures of rRNAs were inferred by alignment to models predicted for published species in the Cimbicidae. For this purpose, the primary and secondary structures of target and references were first aligned in MARNA (Siebert et al., 2005) to identify consensus sequences and consensus structures, respectively. Subsequently, the secondary structures of 12S rRNA and 16S rRNA were predicted by specific structural models in SSU-ALIGN software (Nawrocki et al., 2010). Finally, the structures were artificially transformed into their relative secondary structures with minor changes. The predicted secondary structures of the RNAs were drawn using VARNA (Darty et al., 2009) and RnaViz v. 2.0.3 (De Rijk et al., 2003) programs. The helixes were numbered using the numbering system of *Apis mellifera* (Gillespie et al., 2006) with minor modifications.

2.4 Phylogenetic analyses

PCGs in the mtG were tested for severe substitution saturation with DAMBE (Xia et al., 2003). The unsaturated aligned sequences of the PCGs were concatenated using

chinaXiv:202303.00021v1

SequenceMatrix v. 1.7.8 (Vaidya et al., 2011). The partitioned data block file was used to infer both partitioning schemes and substitution models in PartitionFinder v. 1.1.1 (Lanfear et al., 2012) with "unlinked" branch lengths under the "greedy" search algorithm. The standard partitioning schemes "BIC" and "AICc" were selected for the Bayesian inference (BI) and maximum likelihood (ML) analyses, respectively. Both partitions of the ML and BI analyses selected GTR + I + G as the best-fitting model. The ML analysis was performed using the IQ-TREE web server (https://iqtree.cibiv.univie.ac.at/) (Trifinopoulos et al., 2016) with default parameters except for 0.1 as the perturbation strength and 1,000 as the IQ-TREE stopping rule. The BI analyses were conducted using MrBayes v. 3.2.2 (Ronquist et al., 2012). Four simultaneous Markov chains (three cold and one heated) were run for two million generations in two independent runs with sampling every 1,000 generations and 25% of the first generations discarded as burn-in.

The un-rooted phylogenetic tree under the optimality criterion of ML was inferred using IQ-TREE v. 1.6.10 (Nguyen et al., 2015) with automatically selected best models to align each gene. The alignments from all the genes were analyzed as a single super-matrix to analyze the degree of concatenation. The concatenated file was partitioned based on every gene, and the model for every gene was the automatically selected best model. The node support values were calculated using 1,000 SH-aLRT replicates and 1,000 ultrafast bootstraps (Guindon et al., 2010; Hoang et al., 2018). Individual gene trees for each gene alignment were estimated using IQ-TREE with automatically selected best models and analyzed with ASTRAL-III v. 5.6.1 (Zhang, et al., 2018) to infer the coalescent-based species trees with local branch supports estimated from the quartet frequencies (Sayyari et al., 2016). To reduce the influences of the heterogeneity of the data matrix, we reconstructed the phylogenetic tree using the aa data matrix under the heterogeneous model (LG + C60 + F), as well as selecting the aa with RCFV values < 0.1.

chinaXiv:202303.00021v1

2.5 Estimation of the divergence times

To avoid possible errors caused by the uncertainty of fossil calibration, we performed analyses with three calibration schemes as shown in Table S4 in the PAML package MCMCTree. Prior distribution shapes and scales for calibrations were calculated using the MCMCTreeR package. We used soft maximum bounds with a tail of 0.05 and a prior root for the Cimbicidae at 60-135 Ma. Estimates of the time of divergence time were calculated under Skew-T, Skew-Normal, and Uniform distributions. The analyses were run for 100,000 generations, with a burn-in of 25,000. Every 25th remaining tree was sampled.

chinaXiv:202303.00021v1

3. Results

3.1 Mitogenome architectures

All 17 newly sequenced genomes possess 37 typical genes except for the genome of *Agenocimbex maculatus*, which lacks *trnM* (Fig. S1). Nine genomes failed in the assembly Control Region after multi-strategy assembly was performed. A comparative analysis indicated that the genome organization is highly conserved in the Cimbicidae, including the genome contents, gene order, nucleotide composition, codon usage, and amino acid composition.

The average mitogenome size was 15373 bp with species ranging from 14881 bp for *Corynis lateralis* to 15941 bp for *Palaeocimbex crataegum* (Table S5). AT bias was observed in all the mitogenomes and ranged from 78.47% in *A. maculatus* to 82.68% in *Zaraea akebiae*. The skew metrics were similar across the Cimbicidae mitogenomes with slightly positive A-skewed (0.080 in *Asicimbex concavicaputus* to 0.138 in *Corynis zhengi*) and strongly negative G-skewed (-0.253 in *A. maculatus* to -0.152 in *Z. akebiae*) in the whole mtG.

The bias is correspondingly expressed in the RSCU. The most frequently used codon

was UUA-Leu, which was as high as 5.52 in *A. jimeii*. The RSCU of CCU and CCA, which encode proline, is much larger than those of CCG and CCC, which is consistent with previous studies (Chen, Y. et al., 2020; Doğan et al., 2017; Korkmaz, et al., 2017, 2016, 2015; Niu, et al., 2019; Song, S.-N., Tang et al., 2016; Song, et al., 2016; Tang et al., 2019).

In addition to ATN, TTG appears as the initiation codon in *COX1* of *Abia jimeii* and *ND4* of *Z. mengmeng*. Some genes that encode proteins display lineage preference in the use of termination codons. For example, the Corynidinae evolved to use an incomplete termination codon in *COX1*, while all the rest terminated with a complete one. In addition, the Abiinae use an incomplete termination codon to encode *ND4*. Stable preference in the clades described above could be considered to be an alternative autapomorphy. However, the incomplete termination codon usage of *ND5* seems to be regarded as plesiomorphy of the Cimbicinae because two genera in this subfamily do not have this feature. In addition, this phenomenon is observed in the Corynidinae outside of the subfamily but not in the Abiinae. This indicates that the evolutionary history of molecular derivatives merits more research.

3.2 Gene rearrangements

The gene rearrangements within the Cimbicidae that are the focus of intensive study are the cluster of *CR-trnI-trnQ-trnM-ND2-trnW-trnC-trnY*. By mapping the gene order to the phylogeny based on the mitogenome, we restored the evolution of gene order (Fig. S1). We found that at least the following eight gene rearrangement events could have occurred by deducing the evolutionary history of gene order (Fig. 1):

Type I: The *trnQ-trnM* cluster shuffled and translocated upstream of *trnI*. Although the cluster IQM varies, we reconstructed the ancestral type of Cimbicidae through backtracking as shown in Fig. 1.

Type II: *trnA-trnN-trnR-trnE* is only found in *C. lateralis.*

Type III: The *trnQ-trnM* cluster shuffled and translocated downstream of *trnW*, which is

chinaXiv:202303.00021v1

identical to that in the Abiinae and could be a synapomorphy of this taxon.

Type IV: The patterns of gene rearrangement and rates in the Cimbicinae exhibited more complexity. The translocation of *trnC* and *trnY* occurred in the lineage of Cimbicidae when it diverged from the Abiinae. The resulting cluster of *trnC-trnM-trnQ-trnY,* as a plesiomorphy of clade 5 and 6, is novel to the Hymenoptera.

Type V: A large range of tRNA rearrangement, the *trnQ-trnC-trnW-trnM-trnY-trnI* cluster, has only been observed in clade 4. There may be a mutual cause and effect relationship between the rare event and the uniqueness of the genus *Odontocimbex*.

Type VI: *trnC-trnY* shuffled, translocated, and reversed to upstream of *trnI,* and the *trnM-trnQ* cluster translocated downstream of the control region. The resulting cluster MQ-CR-Y can be regarded as a candidate autapomorphy of clade 1.

Type VII: Three rearrangement events occurred in clade 3. The *trnQ-trnM* cluster and *trnC* translocated upstream of the control region (*trnQ-trnM-trnC-CR*) with *trnY* translocated and reversed upstream of *trnI*.

Type VIII: This is a unique gene order shared by all the *Leptocimbex* species, which were once considered as a synapomorphy of clade 2 and clade 1 (Cheng et al., 2021). They could be revised here as the candidate autapomorphy of the *Leptocimbex*.

3.3 Secondary structure

After accurately predicting the secondary structures of tRNA and rRNA in all the genomes, we conducted a comparative study on these structures. The results show that these conserved secondary structures do contain variations that carry phylogenetic signals.

In the Corynidinae, the stem of H10 in *trnR* is A-U, C-G, and A-U paired, while it is A-U, A-U, A-U, and U-A paired in the Cimbicinae and Abiinae. In clade 3, which consists of *P. tianmunica*, *L. sinicus*, *T. vitellina*, and *A. anthracinum*, the first pair of H10 in *trnF* is G-U, while all the rest are G-C (Figs. S2–S6).

chinaXiv:202303.00021v1

The predicted 16S rRNA length of 25 species is between 1340 and 1377 bp, and they all have six domains and 44 helices (Figs. S7–S11). Domains IV and V were more conserved than the other regions based on a comparison of the secondary structures of 16S rRNA in the Symphyta. In domain I, the Cimbicidae and Abiinae shared a similar secondary structure of H671, which was composed of two short stems and one internal bulge. The first pairing of U-A was replaced by C-G in the Corynidinae. However, some structures were less conserved, such as H235, H589, H837, H1196, and H1648, that varied substantially in the stem-loop structure of the Cimbicidae, which was largely related to the length of stem and the large variation in the number of internal bulges. The third pairing (A-U) in H2547 is the synapomorphy of Abiinae, whereas it is a G-C pairing in the Cimbicinae and Corynidinae.

The predicted 12S rRNA length of 25 species is between 783 and 841 bp, which all have four domains and 26 helices (Figs. S12–16). Domains III and IV were more conserved than those in other regions. Some structures are more stable, such as the stem of H47, the hairpin loop of H673 (GUGAAA), and the stem-loop structure at positions 330-354 of H769, which are stable in the Cimbicidae. The single strand between H921 and H1399 is AAUC in the Corynidinae, AAUU in *C. luteus*, *Palaeocimbex crataegum*, *A. concavicaputus*, and *A. maculatus*, and UAUU in all the other species of Cimbicidae. *P. tianmunica*, *L. sinicus*, *T. vitellina*, and *A. anthracinum* shared the U-G in the sixth pairing of H921, while all the other Cimbicidae species exhibited U-A pairing.

3.4 Low coverage genome (LCG)

On average, Illumina sequencing returned 12.30 (1.67–57.65) GB of sequence data per isolate, which were assembled into ~ 35,282 (6,314–177,647) scaffolds (Table S6). The average depth of the genome coverage was 73.73 × (9.17 ×–349.10 ×). We obtained draft genome assemblies of 119.28–208.29 Mb for the 23 Cimbicids with an average value of N50 of 20.92 (Fig. S17).

chinaXiv:202303.00021v1

3.5 Phylogenetic analyses

The inferred tree topology from the mtG is identical regardless of whether the ML or BI method was used with 95% or higher confidence for most branch nodes (Fig. S18). The topology from OSCs is the same as the former at clade or subfamily levels, except that *Odontocimbex* is a sister group of Clade 1 + 2 + 3 rather than Clade 5 + 6 as inferred from the mtGs.

Five methods with four SCOs data matrices were used, and 19 trees were produced (Fig. S19). Among them, a monophyletic branch of three *Zaraea* species is either nested in the Abiinae or serves as the sister group to all the other members of this subfamily. Species-level relationships within Clade 2 and Clade 5 of the Cimbicinae are not unified. In particular, there are at least two topologies for the relationships among *Leptocimbex nigroteglaris*, *L. yanniae*, and *L. linealis* and *Cimbex luteus Asicimbex concavicaputus* and *Palaeocimbex crataegum*. The three nodes described above also have the lowest support. All three clades contain species that have extremely inadequate BUSCO integrity, or the sample of representative species is too sparse relative to the number of known species. Therefore, we used the Matrix90_Dayhoff6 tree, which is congruent with the morphological view (Yan et al., in preparation) as the primary hypothesis (Fig. 2) for estimating the chronogram.

The basic framework condensed from all the results above is that the Cimbicidae are divided into three major clades (Pachylostictinae is missing). The Cimbicinae form a monophyletic group, which, in turn, forms a sister group with the Abiinae and then aggregates with the Corynidinae. We found six clades in the Cimbicinae. *Zaraea* was recovered to a monophyletic group within the Abiinae. *Abia berezowskii* formed a sister group with *O. sinica* before forming a sister group with (*A. niui* + *A. jimeii*), which is not generally consistent with the studies of (Yan et al., 2020).

3.6 Estimation of divergence times

chinaXiv:202303.00021v1

All the analyses of divergence time recovered largely consistent times of divergence with broadly overlapping credibility intervals regardless of the gene source and prior distribution choice of the fossils (Fig. 2, Fig. S20). From nuclear, we infer an origin of crown Cimbicids in the Upper Cretaceous period (70.8 ~ 82.5 Ma). The Holarctic lineage Cimbicinae + Abiinae originated near the Jurassic-Cretaceous boundary (49.4 ~ 59.2 Ma). The diversification of Cimbicinae occurred at 24.7 ~ 44.5 Ma, and the crown Abiinae emerged at approximately 23.5 ~ 38.4 Ma. The divergence times of each major branch inferred from mtG can basically fall into the above interval. Except for clades related to *Odontocimbex*, because of the incongruence between the mtG and SCG analyses involves the position of *Odontocimbex*.

We selected three fossil calibration points following a thorough literature review and implemented alternative assumptions by manipulating parameterizations of the priors on fossil calibrations confidence intervals for nodes of interest. Whether we regard the oldest fossil as close to the most recent common ancestor (MRCA) of all the extant Cimbicids (Scheme 2) or as close to the MRCA of Cimbicinae + Abiinae (Scheme 3), does not influence inferences on the timing concerning the successful crossing of the K-Pg boundary by the Cimbicidae. Based on biogeographic evidence, the analysis that was performed with scheme 2 provided the best fit, and we provide the results and a discussion of this analysis from this point on (Fig. 2).

4. Discussion

4.1 Mitochondrial genome rearrangement and evolution

It is a common phenomenon that gene rearrangement is found in almost all the sequenced mitogenomes in the Hymenoptera (Aydemir et al., 2020; Cameron, 2014; Chen, Y. et al., 2020; Cheng et al., 2020; Doğan et al., 2017; Dowton, et al., 2009; Korkmaz, Ertan et

chinaXiv:202303.00021v1

al., 2018; Niu, et al., 2019; Yan et al., 2019) Dowton et al. (2009) found that the mtG of the Hymenoptera has a very high rate of gene rearrangements and inferred the Hymenoptera as the basal lineage of the holometabolous orders from comparative analyses. Although only 13 representative species, including three symphytan, were selected in the diversified Hymenoptera, it was concluded after careful deduction that the number and the scale of gene rearrangement were not clearly correlated with lineage. Nevertheless, the conclusions of Song et al. (2016a) could be limited by the steadily increasing number of mtGs in the Symphyta, which could have led to premature conclusions that gene rearrangement is conserved in the Symphyta. On this basis, Ma et al. (Ma et al., 2019a) examined 26 samples from the Symphyta and concluded that gene rearrangement was randomly distributed in the Symphyta. Simultaneously, the authors may be caught up in the overly complex arrangement of genes in the Cephidae and selectively pointed out that gene rearrangement is conserved at the genus level.

In this study, we found abundant gene rearrangements in a family and found that they have some regularity in lineages (Fig. 1). Eight gene rearrangements were found to be shared by 25 species in the Cimbicidae. By mapping gene orders onto the phylogenetic trees derived from the mtG inference, the fit of characters to the phylogeny can be observed. Thus, gene order, as a high-level trait with structural characters, could provide evidence to trace the evolutionary history of some taxa. Three main branching events occurred in the Cimbicidae and could represent changes in the ancestral gene order, which could be the symplesiomorphy of the group. The inversion, transposition, and reverse transposition can be used to explain the mitogenome rearrangements in Cimbicidae. Since the Cimbicidae keep evolving, their patterns of gene rearrangement are becoming increasingly variable. Six subbranch events have been found in the Cimbicidae, which could be the synapomorphy of the group. With an estimation of the evolution history of mitogenome order, the ancestral type of Cimbicidae

chinaXiv:202303.00021v1

was successfully established.

4.2 Comparative analysis of the mitogenome secondary structures

The conserved motifs were identified by comparing them with the mitogenome secondary structures, which can be applied to phylogenetic analyses. The first two pairings (G-U and U-A) of H1792 in the 16S rRNA are unique to the Cimbicidae, while the other reported Tenthredinoidea species to have U-U and C-G pairings (He et al., 2019; Korkmaz et al., 2017; Korkmaz et al., 2016; Korkmaz et al., 2018; Korkmaz et al., 2015; Li et al., 202; Liu et al., 2021; Luo et al., 2019; Ma et al., 2019b; Niu et al., 2019b; Song et al., 2016b; Song et al., 2016a; Wan et al., 2021; Wu, D. et al., 2020; Wu, R. et al., 2019; Yan, 2021). Furthermore, the first pairing at positions 234-335 of H671 in the 16S rRNA is fully conserved in the Abiinae and Cimbicinae, while U-A has been replaced by C-G in *Corynis* and other reported species of Symphyta (except for *Orussus*). Several conserved motifs, such as the first five pairs of the stem-loop structures of H1399, which are fixed in all the species of Symphyta. Compared with the conserved U-A pairings of the other Symphyta, the Cimbicidae shared conserved C-G pairings at positions 325~363 of H769, while the Tenthredinidae shared variable C-G and U-A pairings and exhibited strong phylogenetic signals. The detection of conserved motifs was improved by the use of large numbers of species, which can be used to solve some phylogenetic issues.

4.3 Phylogeny of the Cimbicidae

The three highly supported clades in our results effectively support the subfamilial classification system proposed by Benson (1938), which was later confirmed by Vilhelmsen (2019) using the morphological phylogenetic method. However, previous studies have never been unified in their analyses of the tribes within the subfamily. The interpretation of the phylogeny of the Cimbicidae, whether by Benson or by Vilhelmsen, is primarily based on European fauna and is therefore incomplete. The absence of East Asian genera and species

chinaXiv:202303.00021v1

has led to a bias against a family in which East Asian species account for > 60% (Fig. S19), particularly in the subfamilies Cimbicinae and Abiinae (Yang, et al., 2021). In this study, we found six clades with reliable autapomorphy in the Cimbicinae based on the adequate sampling of the East Asian species. In addition, given the rich diversity of the Abiinae, it is appropriate to recognize the two clades.

Below, we provide diagnoses, content, and comments on the diversity and distribution, diagnosis, and systematics of each major clade.

4.3.1 Subfamily Cimbicinae

Diversity and distribution: Cimbicinae, a Holarctic subfamily including 11 genera and about 116 species, has the highest generic and specific diversity of Cimbicidae. Five genera of the subfamily, *Leptocimbex*, *Odontocimbex*, *Agenocimbex, Asicimbex* and *Labriocimbex*, are endemic to the eastern Asian region. The species diversity of the eastern Asian region is also much higher than other regions of the Holarctic Realm.

Diagnosis (apomorphic characters as AC). Large-sized insects; the anal cell of forewing with a distinct subbasal cross vein, seldom with a punctiformly constricted petiole; clypeus strongly enlarged and distinctly broader than the distance between lower corners of eyes (AC); inner margins of eyes parallel; antenna with 7 or more antennomeres, distinctly longer than head breadth; precoxal bridge present (AC); head strongly enlarged behind eyes in dorsal view (AC); malar space much elongated (AC); posterior margin of first abdominal tergum distinctly incised (AC); female lancet very long and narrow with annuli very short and broad, strongly condensed (AC); apex of lance more or less incised before the short apical process (AC); penis valve with pseudoceps and paravalva distinctly derived (AC); claw enlarged and strongly bent (AC).

Molecular phylogeny. The branching patterns of SCOs and mtG phylogenetic relationships are similar. The order of *Odontocimbex* is slightly different, but it does not

chinaXiv:202303.00021v1

affect the independence of the six internal clades of Cimbicinae.

Clade 1. Clade 1 includes 4 genera and about 54 species (*Trichiosoma* has 34 known species and might be over recognized). Except for the Holarctic *Trichiosoma* and the Palaearctic *Praia*, the other two genera are endemic to Southeastern Asia. Our results showed the sister-group relationship between *Praia* and *Labriocimbex*, which forms a sister group relationship with (*Trichiosoma* + *Asitrichiosoma*). Among the four genera, *Asitrichiosoma* has four known and seven undescribed species. The morphological differences between *Asitrichiosoma* and *Trichiosoma* are distinct and good enough for the generic separation, besides the different distribution patterns. While *Labriocimbex* and *Paria* have species of 5 and 4, respectively.

The synapomorphic characters of the clade 1 are the body very stout and extremely densely pilose; the paravalva of penis valve strongly sclerotized; serrulae very short, broader than long, apex round or truncate.

Despite the rich connotation of clade 1, the monophyly is supported by phylogeny and rare events such as tRNA secondary structure and gene order.

Clade 2. *Leptocimbex*, with 36 valid species and 30 unpublished new species, is probably the most diverse genus in Cimbicidae. Diverse geographical populations and cryptic species have been observed, indicating a comprehensive revision is needed. The species of this clade occur widely in the monsoon region of East Asia.

The synapomorphic characters of the clade are body much elongated; head and thorax with very sparse long hairs; first abdominal tergum with distinct lateral and middle carinae; the paravalva more sclerotized than the pseudoceps and usually with a distinct dorsal corner, and the posterior margin of paravalva usually convex.

Clade 3. *Pseudoclavellaria* has extensive gene rearrangement that forced it to be separated from clade 1.

chinaXiv:202303.00021v1

Pseudoclavellaria occurs broadly in Palaearctic and with only 2 known extant species. *The* synapomorphic characters of the clade are the antenna with five antennomeres; clypeus large, distinctly broadened toward apex; head and thorax very densely pilose; precoxal bridge very broad.

Clade 4. *Odontocimbex* is a monotypic genus and occurs only in the low hills of central China. Differing from other taxa of the subfamily, the larvae of the genus feeding on leaves of *Acanthopanax* species and make hard cocoons on the twigs (Li, T. et al., 2014).

The synapomorphic characters of the clade are: The hind femur with two rows of ventral dents; the mandibles very long and slender, and without inner tooth; head quite small and much narrower than thorax in dorsal view; clypeus strongly extended and longer than broad; cardo and stipes of maxilla long and slender; prementum of labium elongated and about 2 times as long as broad; serrulae long and erect, distinctly narrowed at the middle.

The position of *Odontocimbex* is the only conflict in the phylogenetic relationship constructed by SCOs and mtG. The place of *Odontocimbex* had always been problematic for the similar or unique characters. (Wei et al., 2012) ever briefly discussed the morphological differences between *Odontocimbex* and *Cimbex*, *Palaeocimbex*, and *Trichiosoma*. In this study, *Odontocimbex* is the sister group of (*A. maculatus* + ((*Palaeocimbex crataegum* + *A. concavicaputus*)+ *C. luteus*)), which supported to place *Odontocimbex* to a separate tribe by Deng (2000).

Nevertheless, the position of *Odontocimbex* should be taken with caution and should be followed up in subsequent studies with increasing taxonomic sampling/effective sample size results.

Clade 5. This Clade includes 3 genera and about 20 extant species. Among this group, *Cimbex* is Holarctic and *Palaeocimbex* is Palaearctic. *Asicimbex* is in publishing and includes 9 eastern Asian species (Yang, et al., 2022).

chinaXiv:202303.00021v1

The synapomorphic characters of the clade are: The anal crossvein between 2A and 3A present in a hind wing; lance broad, much broader than lancet and broadened toward subapex; the margins of pseudoceps and paravalva of penis valve roundly curved similarly.

Within this clade, *Palaeocimbex* is sometimes merged with *Cimbex* (Taeger et al., 2010; Vilhelmsen, 2019), but Benson (Benson, 1938), Gussakovskij (1947), Abe and Smith (Abe et al., 1991), Deng (2000), and Wei et al. (Wei et al., 2012)treated it as a valid genus. The morphology of the head, mesoscutellum, and lancet of *Palaeocimbex* are distinct from *Cimbex*. The phylogeny based on SCOs and mtG also supports the validity of *Palaeocimbex*: between *Cimbex* and *Palaeocimbex* lies *Asicimibex*, which is consistent with the previous view (Wei et al., 2012), that the far relationship between *Cimbex* and *Palaeocimbex*. The results of Vilhelmsen (Vilhelmsen, 2019), two *Palaeocimbex* spp. nested with *Cimbex* spp., may be due to the severe shortage of sampling, which leads to the inability to distinguish internal relations of Cimbicinae.

Clade 6. This clade includes 1 genus and 3 species and is endemic to eastern Asia.

The synapomorphic characters of the clade are: the pseudoceps of penis valve strongly protruding; lance very broad and narrowed toward apex; lancet very broad in base and strongly tapering toward apex; serrulae narrow and strongly protruding; mandibles asymmetric, left mandible with three teeth and the right one with two teeth; the first abdominal tergum narrowed toward apex with basal shoulder convex; head strongly narrowed and very short behind eyes in dorsal view (reversed character); malar space very short, about as long as the diameter of median ocellus (a reversed character).

Although *Agenocimbex* and its sister group clade 5 have always formed a monophyly lineage, the evidence from the genomes and also the morphological characters, show that there are significant differences between them. Therefore, Clades 5 and 6 are recognized in this study to show the differences. It is also suggested that *Agenocimbex* should be regarded

chinaXiv:202303.00021v1

as an independent tribe in the practice of taxonomy.

Clade 2+3. The synapomorphic characters of this lineage are: The labrum distinctly broadened toward apex; the serrulae broadened at the middle, and narrowed at base and apex; the clypeus and labrum much paler than other parts of the head.

Clade 1+2+3. The clades 1, 2, and 3 form a distinct lineage among the subfamily Cimbicinae with strong support. The synapomorphic characters of the lineage are: Mandibles distinctly elongated with a narrower base; clypeus strongly broadened and much broader than middle length; the paravalva of penis valve strongly bent at middle.

Clade 1+2+3+4. This clade shares three distinct synapomorphic characters: Claw large, without inner tooth and strongly bent; the pseudoceps of penis valve distinctly protruding at apex; propleura broadly meeting ventrally.

Clades 5+6. This clade is also a distinct lineage among Cimbicinae. The synapomorphic characters of the lineage are: labrum reduced and very small, mostly covered by clypeus; clypeus long and broad, merging with supraclypeal area and distinctly convex anteriorly; claw with the inner tooth strongly appressed to the apical tooth.

4.3.2 Subfamily Abiinae

Diversity and distribution. Abiinae includes 4 genera and about 60 known world species and 30 undescribed eastern Asian species (Yan et al., 2020). The subfamily distributes widely in the Holarctic region, though the most diversities are probably in eastern Asia as two genera and more than 50 species are endemic to this region.

Diagnosis. Medium-sized insects; the inner margins of eyes strongly divergent downwards (AC); dorsal margins of eye closed to each other, especially in males (AC); the lateral furrows of postocellar area absent (AC); frontal walls distinctly elevated (AC); lance with an acute apical process (AC); middle tergites of male usually depressed at middle and with very short and dense hairs (AC); the anal cell in fore wing broadly constricted with a

chinaXiv:202303.00021v1

long middle petiole (AC); clypeus small, much narrower than the lower distance between eyes; labrum large and broad, roundly narrowed toward apex; head weakly dilated behind eyes in dorsal view; claw small, inner tooth present or absent, if present then not appressed to the apical tooth; penis valve simple not differentiated into two lobes.

Molecular phylogeny. Compared with Cimbicinae, Abiinae is less diversified in morphology (Taeger, 1998; Vilhelmsen, 2019). Taeger (1998) synonymized *Zaraea* with *Abia*. Our results showed a clear bifurcation pattern, with *Abia* + *Orientabia* and *Zaraea* recovered as monophyletic. It is consistent with classical morphological classification (Yan et al., 2021). Relationships between *Abia* and *Orientabia* support the previous hypothesis based on morphology (Vilhelmsen, 2015; Hara & Shinohara, 2017; Taeger et al., 2018), that *Abia* is paraphyletic. Although the genera *Abia*, *Orientabia*, and *Allabia* can be easily recognized based on the morphological characters, the validity of the genera *Orientabia* and *Allabia* remains in doubt, which calls for denser sampling to obtain molecular evidence.

4.3.3 Subfamily Corynidinae

Diversity and distribution. This subfamily includes the genus *Corynis* only and about 30 known species. *Corynis* occur mainly within the Mediterranean and Asian desert regions.

Diagnosis. Body small; the inner margins of eyes strongly convergent downwards (AC); head narrowed and very short behind eyes in dorsal view; clypeus and labrum small; maxillary and labial palps strongly reduced (AC); mandibles short and hardly bent; antenna with 5 antennomeres and apical antennomeres strongly enlarged (AC), scape and pedicellum longer than broad; antennal toruli far remote to each other (AC); mesonotum without middle and lateral furrow (AC); the posterior thoracic spiracle concealed; the first abdominal tergum not distinctly incised posteriorly; anal cell in fore wing broadly constricted and with a long middle petiole (AC); hind anal cell with a long petiole (AC); cu-a not interstitial to vein 1M; the inner tibial spur of foreleg much shorter than outer spur (AC); claws small with a large

chinaXiv:202303.00021v1

and separated inner tooth; annuli of lance and lancet not condensed, lance without acute apical process; parapenis large, broadly meeting on meson; penis valve simple.

Molecular phylogeny. The monophyly of Corynidinae was restored in mitochondrial phylogeny. But the gene order of the two species is not the same. There are more than 30 *Corynis* species, which are widely distributed in the Mediterranean, the middle, and west Asian regions. Its evolutionary history and adaptation to the Mediterranean climate still need more representatives to reveal.

4.3.4 Subfamily Pachylostictinae

Diversity and distribution. Five genera and 9 species are included in this subfamily, all occurring in South America.

Diagnosis. Clypeus more or less merged with supraclypeal area (AC) and narrower than the distance between lower corners of eye; labrum broad and roundly narrowed toward apex; antenna with five antennomeres and the second antennomere broader than long (AC); mandibles short and broad, subsymmetrical; deep furrows on mesonotum present; the postspiracle sclerite almost combined to mesepisternum, the suture between them quite fine (AC); mesopleural groove or suture present (AC?); cenchri large and close to each other; vein R+M shorter than the first abscissa of vein 1M; anal cell in fore wing with a broad middle petiole (AC); apex of tibial spur acute; claw small with inner tooth not very close to the apical tooth; lance and lancet not strongly condensed; parapenis long and narrow (?).

Among the five South American genera, *Pachylosticta* and *Pseudopachylosticta* are specialized as shown by the much-reduced mouthparts, venation, and serrulae (Smith, 1988), possibly also the penis valve. They should be sister groups (Vilhelmsen, 2019). *Brasilabia* and *Lopesiana* might be more primitive as they kept some very primitive characters. *Brasilabia* has the first abdominal tergum divided at the middle. *Lopesiana* has the vein cu-a meeting cell 1M at the middle. The hind anal cell is very short and with an extremely long

chinaXiv:202303.00021v1

stalk, as well as the more or less condensed lancet are possibly two synapomorphic characters supporting *Brasilabia*, *Pseudabia*, and *Lopesiana* being monophyletic.

Molecular phylogeny. Pachylostictinae was not included in this study because of a lack of fresh materials. The monophyly of the subfamily and its systematic position within Cimbicidae can not be verified at present. From the present state of diversity, distribution pattern, and external morphology, it might be a monophyletic group and an ancient clade of the family (Vilhelmsen, 2019). The ancient characters will make it difficult to encode in the matrix, which leads to the dilemma that the monophyly could not be recovered in morphological phylogenetic inference.

In contrast to the other three subfamilies, the monophyly of the Pachylostictinae was not always corroborated, but putative autapomorphies were identified. Despite having the lowest species diversity among the cimbicid subfamilies, the Pachylostictinae are morphologically diverse at the genus level, indicating that they have evolved in isolation for a long time (Vilhelmsen, 2019).

4.3.5 Abiinae + Cimbicinae

The synapomorphic characters of the lineage are: Parapenis strongly reduced and unrecognizable; apex of the lance with an acute process; lance and lancet strongly condensed with more than 30 very short annuli; apex of tibial spur membranous.

4.3.6 Corynidinae + Pachylostictinae

The synapomorphic characters of the lineage are: Antenna with five antennomeres and the last antennomeres enlarged or elongated; anal cell in fore wing with a long middle petiole; female eyes distinctly convergent downwards.

4.4 Divergence

Our divergence times are notably younger than those found in other studies (Table S2) and particularly younger than those in the mitogenome phylogeny (Niu et al., 2021). One

chinaXiv:202303.00021v1

reason could be the sampling. Obtaining the reliable crown age requires sampling from the member derived from the deepest split. The insufficient knowledge of interfamilial phylogenetic relationships and unavailability of molecular markers for the Pachylostictinae may cause the crown to age younger (Brown et al., 2018).

The oldest fossil of the Cimbicidae, *Cenocimbex menatensis*, that was reported from Menat, France, and dated to 59 Ma was revised to 60-61 Ma by Schubnel et al. (2020). Owing to its morphological similarity with the modern Holarctic species, we hypothesized that the Cimbicidae should have differentiated earlier. However, there are no fossils in this period close to the abundance of forms described from 30 and 20 strata, which could suggest that early evolution in the Cimbicidae did not rapidly diversify, and the subfamilies Cimbicinae and Abiinae diversified explosively in the Cenozoic period through vicariance.

The chronograms are consistent with the inference from fossils. The ensuing mass extinction devastated the global vegetation, but after 0.5 Myr, the vegetation gradually recovered. The surviving Cimbicidae had the opportunity to diversify. With the rebounding of angiosperm species, they radiated to the Holarctic and spread to Gondwana through the island channel between North and South America in the early Cenozoic period.

The living species of the Pachylostictinae subfamily in Gondwana are very limited in both genera and species. However, the families related to the Cimbicidae, such as the Argidae, have 3-4 subfamilies in South America and high generic and specific diversity. Simultaneously, the Selandriinae, probably the first transoceanic traveler of the Tenthredinidae, also has high generic diversity in South America. In contrast, the Pachylostictinae subfamily could be a group with inadequate potential for diversity.

However, even so, the Pachylostictinae subfamily in this habitat, like mammals after the Cretaceous period, adapted various types of morphology that enabled it to expand its niche rapidly. Because of the lack of fossils, it is not known whether all these attempts to diversify

chinaXiv:202303.00021v1

their forms have been successful. At least for now, there are five living genera in the Pachylostictinae subfamily, which are scattered and may have faced many extinction events.

When the Pachylostictinae diffused through the pore (or the South American continent was isolated from it after independence), *Corynis*, which is considered to be the sister group of the former, remained in Laurasia. However, the character in the middle that broadly constricted the anal cell in the forewing shared by the extant *Corynis* and Abiinae is only found in the fossils from Shanwang. None of these fossils have been assigned to *Corynis*. As a result, the fossil record of *Corynis* is still blank. Considering its early origin and wide distribution, *Corynis*, which only contains 30 species, cannot be regarded as a diverse lineage. This is consistent with the characteristics that the diversified groups usually have a low-speed basal branch. This subfamily is confined to the region dominated by the Mediterranean climate, including the Mediterranean Sea and the desert areas of Southwest Asia, West Asia and Central Asia, Xinjiang and western Inner Mongolia in China, and few species can extend to northern Europe. The mtG time tree supports their differentiation from 33.2 Ma, and more samples may make this time younger. In general, the evolutionary trajectory of *Corynis* has probably been impacted by the primary paleo-climatic events that have occurred in the Mediterranean Basin since the Miocene. This is consistent with a recent review that shows phylogenetic evidence for a Miocene origin of various Mediterranean flora with different life forms and biogeographic histories (Vargas et al., 2018). However, how the Messinian Salinity Crisis (MSC: 5.96 ± 0.02 Ma; Krijgsman et al., 1999), the onset of the Mediterranean climate (3.4-2.8 Ma; Suc 1984), and the subsequent climatic oscillations in the Pleistocene (Kadereit et al., 2004) shaped the diversity of *Corynis* require deduction by dense sampling.

Compared with the Pachylostictinae in Gondwana, the Cimbicidae on the ancient land of Laurasia had experienced relatively far-reaching succession. Shortly after the mass extinction (63.4–56 Ma), there was a rapid dispersal and divergence. The fossils in Okanagan Highland

chinaXiv:202303.00021v1

(47–56 Ma) indicate that the Cimbicidae at that time were relatively modern. However, the diversity was fully established by the Neogene period when diverse assemblages also appeared in Shanwang at that time.

The diversity of crown Cimbicinae could have occurred within 43.4 Ma. Dense fossil sampling in the Ypresian and Burdigalian regions suggests that the family suddenly diversified, which is consistent with the time of divergence of the major clades that were deduced. The diversity of genera had almost been completed at approximately 10 Ma. It is hypothesized that they survived the Oligocene glacial period when the broad-leaved forests declined (Meng, et al., 1998) and spread to Europe around the Grande Coupure mass extinction (33.5 Ma). Simultaneously, the uplift of the Himalayas promoted the formation of an East Asian monsoon climate and provided the warm and humid environment (Deng, Tao et al., 2021) that helped to enable the continuous diversification of the East Asian flora (Chen, et al., 2018) and fauna, shaping the same biogeographic pattern as that found in the Abiinae.

The absence of Paleogene Abiinae fossils could indicate that the initial diversification only occurred in the Neogene period. This is consistent with our deduction that the time of divergence was 38.4 Ma. The establishment of the Bering land bridges around the Palaeogene/Neogene boundary (Marincovich et al., 2001) could have enabled the subsequent colonization of North American from Eurasia. The closure of the only passage ended the faunal exchange between Eurasia and North American. The Abiinis evolved independently and formed their current distribution pattern that is dominated by Eurasia, with only four species of two genera distributed in North America. There are no endemic genera but a few endemic species. However, owing to the lack of American samples, we cannot rule out the possibility that the transoceanic diffusion occurred in the Pliocene closure (< 5 Ma).

4.5 Geographical distribution pattern

The distributions of invasive species are not included in the following discussion.

chinaXiv:202303.00021v1

The distribution of Cimbicidae is unique. The family is predominantly distributed in the northern hemisphere with the exception of a few clades. This differs markedly with the following families of the basal lineages of Hymenoptera. The extant Anaxyelidae are only distributed in North America, though the fossil taxa are more diversified in different areas of world. The Blasticotomidae and Heptamelidae are distributed in Eurasia, but their diversity is concentrated in East Asia. The Megalodontesidae are distributed in the Palaearctic realm, but their diversity is concentrated in the Mediterranean region. The Pergidae are distributed in Australia and the neotropical realm, while the Athaliidae are distributed in Eurasia and Africa. The Orussidae are distributed globally, but their diversity in the Southern Hemisphere is higher than that of the Northern Hemisphere. The Xiphydriidae are distributed globally, but their diversity is concentrated in the southern part of Asia. The Argidae are distributed globally, but they have three major centers of diversity in South America, East Asia and Africa (Malagón-Aldana et al., 2021). The characteristics of the distribution of all these groups differ distinctly from those of the Cimbicidae.

Several families are distributed in a manner similar to that of the Cimbicidae. They are primarily Holarctic with diversity in East Asia. However, they differ in their local distribution at higher resolution. The Cephidae are widely distributed throughout the Holarctic realm, but their diversity at the genus level is primarily in East Asia, and the species level of diversity is relatively similar in southern Europe and eastern Asia. However, there is one endemic genus and one to two endemic species each in Madagascar, Australia, and Indonesia. The Siricidae are primarily distributed in Eurasia, but there is one endemic genus and six endemic species in Africa and two endemic genera in Central America. One is in Mexico, and the other is in Cuba. Compared with them, most of the clades of Cimbicidae are distributed throughout the Holarctic realm. Two extant subfamilies (Cimbicinae and Abiinae) are primarily distributed in East Asia; one subfamily (Corynidinae, only one genus) is distributed in the Mediterranean

chinaXiv:202303.00021v1

region, and the subfamily Pachylostictinae, which harbors several primitive morphological traits, is entirely distributed in the central part of the South America, far from the main distribution of the family.

Above all, at the family level, this distribution of Cimbicidae differs significantly from all the other families of the nasal lineages of Hymenoptera. What triggered this distribution and the key reason for it require a specific interpretation. If the hierarchy of the group is omitted, there are six subfamilies of Tenthredinidae, including Selandriinae, Fenusinae, Belesinae, Lycaotinae, Blennocampinae, and Allantinae, that are distributed in a manner similar to that of the Cimbicidae. They are dominant in Eurasia with a small number of endemic species distributed in the hinterland of South America, though the detailed characteristics of the distribution of these subfamilies still differ from those of the Cimbicidae. However, a comparison of the family with the subfamilies requires either ignoring evolutionary hierarchies as the prerequisite or reconsidering the relationship between geographical distribution patterns and taxon rank.

Given the possible causes of the biogeographic patterns of these groups, the history of divergence of the early group and continental drift processes of these groups merits intensive study. We carefully deduced from the existing evidence that the process of evolution of the Cimbicidae might have the following features. 1) Around the Middle Cretaceous period (approximately 120–100 Myr), the early clades of these groups were relatively concentrated in some regions of Eurasia. 2) Around 80 Myr, some groups or offspring crossed through North Africa into the South America hinterland by land connections. 3) After the separation in South America, a small number of these groups had a high potential for differentiation (Wei et al., 2010) and gradually diversified in South America. However, the vast majority of groups do not have the potential to differentiate rapidly, and in the new environment of South America, most gradually became extinct. A few groups survive to this day, resulting in a low

chinaXiv:202303.00021v1

level of species diversity in each genus. 4) During the same period, the subject group retained in Eurasia rapidly diversified in East Asia or the Mediterranean region and formed the current group subject. 5) The main group in Eurasia, which entered North America in batches at different periods, formed the Tenthredinidae fauna in North America, and a few of these groups expanded from North America to South America approximately 50 Myr by moving into Central America. Owing to the completely different climatic and natural geographical conditions of Central and South America, these groups stopped at Central America and differentiated into a rare endemic species in this region.

The biogeographic pattern formation process of the southern bound group of the Pergidae and Argidae may differ significantly from this. There are three differences. 1) The two families originated much earlier, and they had probably originated and began to diverge at approximately 160 Myr in the relatively early stages of the Pangaea division. 2) The distribution region subjects of the early sublines of these groups on Pangea may be larger or more biased towards Africa and South America and even cover the major regions of Africa and South America. After the combined Pangea split and gradually drifted, these groups retained more groups in South America and Africa. 3) On the South American and African mainland, the major early sublines of these retained groups certainly contain more groups that could have diversified. They formed the more current and prominent diversity of the Tenthredinidae of South America and Africa, which contain not only many large genera with rich species but also more monotypic or oligotypic genera.

5. Conclusions

A robust phylogenetic relationship is a prerequisite to deducing evolutionary history. Combining molecular sequence data with fossil records is expected to restore the influence of Earth history on the cladistic evolution history of extant fauna.

chinaXiv:202303.00021v1

In this study, the phylogenetic relationship of Cimbicidae was constructed using dense samples and comprehensive genome sequences. Gene rearrangement and secondary structure were used as molecular sources to strengthen the understanding of the evolutionary history of the Cimbicidae. The use of a reasonable fossil calibration strategy provides the chronogram as the framework to combine with the evolution of key morphological features. Finally, we addressed the classification of Cimbicidae. We concluded that the connotations of the three Holarctic subfamilies should be updated. Among them, six clades can be distinguished from the subfamily Cimbicinae, and two clades have been identified in the Abiinae. Combining the evidence from the phylogeny of extant species and the diversity and patterns of distribution of the Cimbicidae, we attempted to restore the evolutionary history of this family.

Although the increase in Cimbicidae phylogenetic diversity arose early, before the K-Pg boundary, their diversity has arisen since the Neogene period. The complexity was established by a pulse rather than by a continuous process of diversification. This study provides a model to explore the rationality of the existence of high ranks and the reasons for their formation. In addition, it urges us to reflect on the role and weight of the evolution of angiosperms in the genesis of insect diversity.

chinaXiv:202303.00021v1

ACKNOWLEDGEMENTS

We would like to thank Xiao Wei, and Li Zejian for their assistance in sample collection. We are grateful to the anonymous reviewers of the manuscripts. This work was supported by the National Natural Science Foundation of China (Grant No. 31970447).

References

1 Abe, M., & Smith, D.R. The Genus-group Names of Symphyta (Hymenoptera) and Their Type Species. Esakia, 1991, 31115.

2 Aydemir, M.N., & Korkmaz, E.M. Comparative mitogenomics of Hymenoptera reveals evolutionary differences in structure and composition. International Journal of Biological Macromolecules, 2020, 144460–472. Retrieved December 26, 2021, from https://linkinghub.elsevier.com/retrieve/pii/S0141813019349955. DOI: 10.1016/j.ijbiomac.2019.12.135

3 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 2012, 19(5): 455–477. Retrieved February 28, 2023, from http://www.liebertpub.com/doi/10.1089/cmb.2012.0021. DOI: 10.1089/cmb.2012.0021

4 Benson, R.B. On the classification of sawflies (Hymenoptera Symphyta). Transactions of the Royal Entomological Society of London, 1938, 87(15): 353–384. Retrieved December 26, 2021, from https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2311.1938.tb00721.x. DOI: 10.1111/j.1365-2311.1938.tb00721.x

5 Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution, Mitogenomics and Metazoan Evolution , 2013, 69(2): 313–319. Retrieved December 26, 2021, from https://www.sciencedirect.com/science/article/pii/S1055790312003326. DOI: 10.1016/j.ympev.2012.08.023

6 Brown, J.W., & Smith, S.A. The Past Sure is Tense: On Interpreting Phylogenetic Divergence Time Estimates. Systematic Biology, 2018, 67(2): 340–353. Retrieved December 26, 2021, from https://doi.org/10.1093/sysbio/syx074. DOI: 10.1093/sysbio/syx074

7 Bushnell, B. BBTools Software Package. Available online: https://sourceforge.net/projects/bbmap/ (accessed on 24 October 2020), 2020.

8 Cameron, S.L. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. Annual Review of Entomology, 2014, 59(1): 95–117. Retrieved May 8, 2022, from https://www.annualreviews.org/doi/10.1146/annurev-ento-011613-162007. DOI: 10.1146/annurev-ento-011613-162007

9 Capella-Gutiérrez, S., Silla-Martínez, J.M., & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 2009, 25(15): 1972–1973. Retrieved February 28, 2023, from https://academic.oup.com/bioinformatics/article/25/15/1972/213148. DOI: 10.1093/bioinformatics/btp348

10 Chen, Y., Wei, M., Yang, H., Wang, H., & Niu, G. Nearly complete mitochondrial genome of *Trichiosoma vitellina* Linné, 1760 (Hymenoptera: Tenthredinidae): sequencing and phylogenetic analysis. Mitochondrial DNA Part B, 2020, 5(1): 802–803. Retrieved December 8, 2021, from https://www.tandfonline.com/doi/full/10.1080/23802359.2020.1715860. DOI: 10.1080/23802359.2020.1715860

11 Chen, Y.-C., Wang, C.-T., Lees, D.C., & Wu, L.-W. Higher DNA insert fragment sizes improve mitogenomic assemblies from metagenomic pyrosequencing datasets: an example using Limenitidinae butterflies (Lepidoptera, Nymphalidae). Mitochondrial DNA Part A, 2018, 29(6): 840–845. Retrieved November 22, 2021, from https://www.tandfonline.com/doi/full/10.1080/24701394.2017.1373106. DOI:

chinaXiv:202303.00021v1

10.1080/24701394.2017.1373106
12 Cheng, Y., Yan, Y., Wei, M., & Niu, G. The complete mitochondrial genome of *Praia tianmunica* (Hymenoptera: Cimbicidae) with related phylogenetic analysis. Mitochondrial DNA Part B, 2020, 5(3): 3037–3038. Retrieved December 8, 2021, from https://www.tandfonline.com/doi/full/10.1080/23802359.2020.1797576. DOI: 10.1080/23802359.2020.1797576
13 Cheng, Y., Yan, Y., Wei, M., & Niu, G. Characterization of mitochondrial genomes of three new species: *Leptocimbex praiaformis* , *L. CLAVICORNIS* , and *L. yanniae* (Hymenoptera: Cimbicidae). Entomological Research, 2021, 51(6): 287–304. Retrieved February 28, 2023, from https://onlinelibrary.wiley.com/doi/10.1111/1748-5967.12501. DOI: 10.1111/1748-5967.12501
14 Chikhi, R., & Rizk, G. Space-Efficient and Exact de Bruijn Graph Representation Based on a Bloom Filter. In B. Raphael & J. Tang (Eds.), Algorithms in Bioinformatics, Lecture Notes in Computer Science (Vol. 7534, pp. 236–248). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved February 27, 2023, from http://link.springer.com/10.1007/978-3-642-33122-0_19. DOI: 10.1007/978-3-642-33122-0_19
15 Darty, K., Denise, A., & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics, 2009, 25(15): 1974–1975. Retrieved February 28, 2023, from https://academic.oup.com/bioinformatics/article/25/15/1974/210730. DOI: 10.1093/bioinformatics/btp250
16 De Rijk, P., Wuyts, J., & De Wachter, R. RnaViz 2: an improved representation of RNA secondary structure. Bioinformatics, 2003, 19(2): 299–300. Retrieved February 28, 2023, from https://academic.oup.com/bioinformatics/article/19/2/299/372776. DOI: 10.1093/bioinformatics/19.2.299
17 Deng, Tao, Wu, F., Wang, S., Su, T., & Zhou, Z. Major turnover of biotas across the Oligocene/Miocene boundary on the Tibetan Plateau. Palaeogeography, Palaeoclimatology, Palaeoecology, 2021, 567110241. Retrieved December 23, 2021, from https://www.sciencedirect.com/science/article/pii/S0031018221000262. DOI: 10.1016/j.palaeo.2021.110241
18 Deng, Tiejun *A Systematic Study of the Cimbicidae from China (Hymenoptera, Symphyta)* (Master's Thesis). Central South Forestry University.
19 Doğan, Ö., & Korkmaz, E.M. Nearly complete mitogenome of hairy sawfly, Corynis lateralis (Brullé, 1832) (Hymenoptera: Cimbicidae): rearrangements in the IQM and ARNS1EF gene clusters. Genetica, 2017, 145(4–5): 341–350.
20 Dowton, M., Cameron, S.L., Austin, A.D., & Whiting, M.F. Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – A lineage with both rapidly and slowly evolving mitochondrial genomes. Molecular Phylogenetics and Evolution, 2009, 52(2): 512–519. Retrieved April 18, 2022, from https://www.sciencedirect.com/science/article/pii/S1055790309001262. DOI: 10.1016/j.ympev.2009.04.001
21 Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., & Whiting, M.F. Characterization of 67 Mitochondrial tRNA Gene Rearrangements in the Hymenoptera Suggests That Mitochondrial tRNA Gene Position Is Selectively Neutral. Molecular Biology and Evolution, 2009, 26(7): 1607–1617. Retrieved April 17, 2022, from https://doi.org/10.1093/molbev/msp072. DOI: 10.1093/molbev/msp072
22 Gillespie, J.J., Johnston, J.S., Cannone, J.J., & Gutell, R.R. Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of Apis mellifera (Insecta: Hymenoptera): structure, organization, and retrotransposable elements. Insect Molecular Biology, 2006, 15(5): 657–686. Retrieved September 18, 2022, from

chinaXiv:202303.00021v1

https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2583.2006.00689.x. DOI: 10.1111/j.1365-2583.2006.00689.x
23 Goulet, H. *The genera and subgenera of the sawflies of Canada and Alaska: Hymenoptera:Symphyta*. The Insects and arachnids of Canada . Canada: Research Branch, Agriculture Canada.
24 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology, 2010, 59(3): 307–321. Retrieved February 28, 2023, from https://academic.oup.com/sysbio/article/59/3/307/1702850. DOI: 10.1093/sysbio/syq010
25 He, H., Niu, G., Zhang, B., & Wei, M. The complete mitochondrial genome of *Athalia proxima* (Hymenoptera: Tenthredinidae) and phylogenetic analysis. Mitochondrial DNA Part B, 2019, 4(2): 3868–3869. Retrieved February 28, 2023, from https://www.tandfonline.com/doi/full/10.1080/23802359.2019.1687042. DOI: 10.1080/23802359.2019.1687042
26 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., & Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution, 2018, 35(2): 518–522. Retrieved February 28, 2023, from https://academic.oup.com/mbe/article/35/2/518/4565479. DOI: 10.1093/molbev/msx281
27 Isaka, Y., & Sato, T. Was species diversification in Tenthredinoidea (Hymenoptera: Symphyta) related to the origin and diversification of angiosperms? The Canadian Entomologist, 2015, 147(4): 443–458. Retrieved November 23, 2021, from https://www.cambridge.org/core/product/identifier/S0008347X14000601/type/journal_article. DOI: 10.4039/tce.2014.60
28 Kadereit, J.W., Griebeler, E.M., & Comes, H.P. Quaternary diversification in European alpine plants: pattern and process. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 2004, 359(1442): 265–274.
29 Katoh, K., & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics, 2008, 9(4): 286–298. Retrieved February 28, 2023, from https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbn013. DOI: 10.1093/bib/bbn013
30 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 2012, 28(12): 1647–1649. Retrieved February 27, 2023, from https://academic.oup.com/bioinformatics/article/28/12/1647/267326. DOI: 10.1093/bioinformatics/bts199
31 Korkmaz, E. Mahir, Aydemir, H.B., Temel, B., Budak, M., & Basibuyuk, H.H. Mitogenome evolution in Cephini (Hymenoptera: Cephidae): Evidence for parallel adaptive evolution. Biochemical Systematics and Ecology, 2017, 71137–146. Oxford: Pergamon-Elsevier Science Ltd. Retrieved March 8, 2022, from https://www.webofscience.com/wos/alldb/full-record/WOS:000401046000017. DOI: 10.1016/j.bse.2017.02.004
32 Korkmaz, E. Mahir, Aydemir, H.B., Temel, B., Budak, M., & Başıbüyük, H.H. Mitogenome evolution in Cephini (Hymenoptera: Cephidae): Evidence for parallel adaptive evolution. Biochemical Systematics and Ecology, 2017, 71137–146. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S0305197817300285. DOI: 10.1016/j.bse.2017.02.004
33 Korkmaz, E. Mahir, Budak, M., Ördek, M.N., & Başıbüyük, H.H. The complete mitogenomes of Calameuta filiformis (Eversmann, 1847) and Calameuta idolon (Rossi,

chinaXiv:202303.00021v1

1794) (Hymenoptera: Cephidae): The remarkable features of the elongated A+T rich region in Cephini. Gene, 2016, 576(1): 404–411. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S0378111915012585. DOI: 10.1016/j.gene.2015.10.050

34 Korkmaz, Ertan Mahir, Doğan, Ö., Budak, M., & Başıbüyük, H.H. Two nearly complete mitogenomes of wheat stem borers, Cephus pygmeus (L.) and Cephus sareptanus Dovnar-Zapolskij (Hymenoptera: Cephidae): An unusual elongation of rrnS gene. Gene, 2015, 558(2): 254–264. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S0378111915000074. DOI: 10.1016/j.gene.2014.12.069

35 Korkmaz, Ertan Mahir, Dogan, O., Durel, B.S., Altun, B.T., Budak, M., & Basibuyuk, H.H. Mitogenome organization and evolutionary history of the subfamily Cephinae (Hymenoptera: Cephidae). Systematic Entomology, 2018, 43(3): 606–618. Hoboken: Wiley. Retrieved March 8, 2022, from https://www.webofscience.com/wos/alldb/full-record/WOS:000434280000011. DOI: 10.1111/syen.12290

36 Korkmaz, Ertan MahiR, Doğan, Ö., Durel, B.S., Temel Altun, B., Budak, M., & Başibüyük, H.H. Mitogenome organization and evolutionary history of the subfamily Cephinae (Hymenoptera: Cephidae): Divergent selection in Cephinae mitogenomes. Systematic Entomology, 2018, 43(3): 606–618. Retrieved February 28, 2023, from https://onlinelibrary.wiley.com/doi/10.1111/syen.12290. DOI: 10.1111/syen.12290

37 Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., & Wilson, D.S. Chronology, causes and progression of the Messinian salinity crisis. Nature, 1999, 400(6745): 652–655. Retrieved December 26, 2021, from https://www.nature.com/articles/23231/. DOI: 10.1038/23231

38 Kück, P., & Struck, T.H. BaCoCa – A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Molecular Phylogenetics and Evolution, 2014, 7094–98. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S1055790313003655. DOI: 10.1016/j.ympev.2013.09.011

39 Kumar, S., Stecher, G., & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution, 2016, 33(7): 1870–1874. Retrieved February 27, 2023, from https://academic.oup.com/mbe/article/33/7/1870/2579089. DOI: 10.1093/molbev/msw054

40 Lanfear, R., Calcott, B., Ho, S.Y.W., & Guindon, S. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Molecular Biology and Evolution, 2012, 29(6): 1695–1701. Retrieved February 28, 2023, from https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss020. DOI: 10.1093/molbev/mss020

41 Li, T., Hua, J., Wright, A.M., Cui, Y., Xie, Q., Bu, W., & Hillis, D.M. Long-branch attraction and the phylogeny of true water bugs (Hemiptera: Nepomorpha) as estimated from mitochondrial genomes. BMC Evolutionary Biology, 2014, 14(1): 99. Retrieved November 21, 2021, from https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-14-99. DOI: 10.1186/1471-2148-14-99

42 Li, Y., Wei, M., Liu, J., & Niu, G. Characterization of the mitochondrial genome of *Eutomostethus vegetus* Konow, 1898 (Hymenoptera: Tenthredinidae) and phylogenetic analysis. Mitochondrial DNA Part B, 2020, 5(3): 3033–3034. Retrieved February 28, 2023, from https://www.tandfonline.com/doi/full/10.1080/23802359.2020.1797563. DOI: 10.1080/23802359.2020.1797563

43 Liu, Y., Wei, M., & Niu, G. The first mitochondrial genome of a fern sawfly, *Strongylogaster xanthocera* Stephens, 1835 (Hymenoptera: Tenthredinidae).

chinaXiv:202303.00021v1

Mitochondrial DNA Part B, 2021, 6(3): 902–904. Retrieved February 28, 2023, from https://www.tandfonline.com/doi/full/10.1080/23802359.2021.1886019. DOI: 10.1080/23802359.2021.1886019

44 Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience, 2012, 1(1): 18. Retrieved February 28, 2023, from https://academic.oup.com/gigascience/article-lookup/doi/10.1186/2047-217X-1-18. DOI: 10.1186/2047-217X-1-18

45 Luo, X., Wei, M., & Niu, G. Nearly complete mitochondrial genome of *Siobla xizangensis* Xiao, Huang & Zhou, 1988 (Hymenoptera: Tenthredinidae) and phylogenetic analysis. Mitochondrial DNA Part B, 2019, 4(2): 4102–4103. Retrieved February 28, 2023, from https://www.tandfonline.com/doi/full/10.1080/23802359.2019.1692729. DOI: 10.1080/23802359.2019.1692729

46 Ma, Y., Zheng, B., Zhu, J., van Achterberg, C., Tang, P., & Chen, X. The first two mitochondrial genomes of wood wasps (Hymenoptera: Symphyta): Novel gene rearrangements and higher-level phylogeny of the basal hymenopterans. International Journal of Biological Macromolecules, 2019 a, 1231189–1196. Retrieved April 19, 2022, from https://linkinghub.elsevier.com/retrieve/pii/S0141813018350141. DOI: 10.1016/j.ijbiomac.2018.11.017

47 Ma, Y., Zheng, B., Zhu, J., van Achterberg, C., Tang, P., & Chen, X. The first two mitochondrial genomes of wood wasps (Hymenoptera: Symphyta): Novel gene rearrangements and higher-level phylogeny of the basal hymenopterans. International Journal of Biological Macromolecules, 2019 b, 1231189–1196. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S0141813018350141. DOI: 10.1016/j.ijbiomac.2018.11.017

48 Malagón-Aldana, L.A., Smith, D.R., Shinohara, A., & Vilhelmsen, L. From Arge to Zenarge: adult morphology and phylogenetics of argid sawflies (Hymenoptera: Argidae). Zoological Journal of the Linnean Society, 2021, 193(3): 880–938. Retrieved December 26, 2021, from https://doi.org/10.1093/zoolinnean/zlaa170. DOI: 10.1093/zoolinnean/zlaa170

49 Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., & Zdobnov, E.M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. (J. Kelley, Ed.) Molecular Biology and Evolution, 2021, 38(10): 4647–4654. Retrieved January 12, 2022, from https://academic.oup.com/mbe/article/38/10/4647/6329644. DOI: 10.1093/molbev/msab199

50 Marincovich, L., & Gladenkov, A.Y. New evidence for the age of Bering Strait. Quaternary Science Reviews, 2001, 20(1–3): 329–335. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S027737910000113X. DOI: 10.1016/S0277-3791(00)00113-X

51 Meng, G., Li, Y., Yang, C., & Liu, S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Research, 2019, 47(11): e63.

52 Meng, J., & McKenna, M.C. Faunal turnovers of Palaeogene mammals from the Mongolian Plateau. Nature, 1998, 394(6691): 364–367. Nature Publishing Group. Retrieved March 2, 2022, from https://www.nature.com/articles/28603. DOI: 10.1038/28603

53 Nawrocki, E.P., & Farm, H.J. SSU-ALIGN User's Guide, 2010, .

54 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., & Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution, 2015, 32(1): 268–274. Retrieved November 18, 2022,

chinaXiv:202303.00021v1

from https://doi.org/10.1093/molbev/msu300. DOI: 10.1093/molbev/msu300
55 Niu, G., Jiang, S., Doğan, Ö., Korkmaz, E.M., Budak, M., Wu, D., & Wei, M. Mitochondrial Phylogenomics of Tenthredinidae (Hymenoptera: Tenthredinoidea) Supports the Monophyly of Megabelesesinae as a Subfamily. Insects, 2021, 12(6): 495. Retrieved December 8, 2021, from https://www.mdpi.com/2075-4450/12/6/495. DOI: 10.3390/insects12060495
56 Niu, G., Korkmaz, E.M., Doğan, Ö., Zhang, Y., Aydemir, M.N., Budak, M., Du, S., et al. The first mitogenomes of the superfamily Pamphilioidea (Hymenoptera: Symphyta): Mitogenome architecture and phylogenetic inference. International Journal of Biological Macromolecules, 2019 a, 124185–199. Retrieved December 8, 2021, from https://linkinghub.elsevier.com/retrieve/pii/S0141813018346762. DOI: 10.1016/j.ijbiomac.2018.11.129
57 Niu, G., Korkmaz, E.M., Doğan, Ö., Zhang, Y., Aydemir, M.N., Budak, M., Du, S., et al. The first mitogenomes of the superfamily Pamphilioidea (Hymenoptera: Symphyta): Mitogenome architecture and phylogenetic inference. International Journal of Biological Macromolecules, 2019 b, 124185–199. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S0141813018346762. DOI: 10.1016/j.ijbiomac.2018.11.129
58 Niu, G., Zhang, Y., Li, Z., & Wei, M. Characterization of the mitochondrial genome of *Analcellicampa xanthosoma* gen. et sp. nov. (Hymenoptera: Tenthredinidae). PeerJ, 2019, 7e6866. Retrieved December 8, 2021, from https://peerj.com/articles/6866. DOI: 10.7717/peerj.6866
59 Nyman, T., Onstein, R.E., Silvestro, D., Wutke, S., Taeger, A., Wahlberg, N., Blank, S.M., et al. The early wasp plucks the flower: disparate extant diversity of sawfly superfamilies (Hymenoptera: 'Symphyta') may reflect asynchronous switching to angiosperm hosts. Biological Journal of the Linnean Society, 2019, 128(1): 1–19. Retrieved November 23, 2021, from https://academic.oup.com/biolinnean/article/128/1/1/5526148. DOI: 10.1093/biolinnean/blz071
60 O'Reilly, J.E., dos Reis, M., & Donoghue, P.C.J. Dating Tips for Divergence-Time Estimation. Trends in Genetics, 2015, 31(11): 637–650. Retrieved December 26, 2021, from https://www.sciencedirect.com/science/article/pii/S0168952515001468. DOI: 10.1016/j.tig.2015.08.001
61 Perna, N.T., & Kocher, T.D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution, 1995, 41(3): 353–358. Retrieved December 26, 2021, from http://link.springer.com/10.1007/BF01215182. DOI: 10.1007/BF01215182
62 Pryszcz, L.P., & Gabaldón, T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Research, 2016, 44(12): e113–e113. Retrieved February 27, 2023, from https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw294. DOI: 10.1093/nar/gkw294
63 Ren X., Wang M., Wang H., & Li G. Progress in the research of taxonomic studies on Cimbicidae (Insecta: Hymenoptera). Forest Pest and Disease, 2020, 39(5): 36–48.
64 Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L., & Rasnitsyn, A.P. A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. Systematic Biology, 2012, 61(6): 973–999. Retrieved November 23, 2021, from https://academic.oup.com/sysbio/article/61/6/973/1665823. DOI: 10.1093/sysbio/sys058
65 Sahlin, K., Vezzi, F., Nystedt, B., Lundeberg, J., & Arvestad, L. BESST - Efficient scaffolding of large fragmented assemblies. BMC Bioinformatics, 2014, 15(1): 281.

chinaXiv:202303.00021v1

Retrieved February 28, 2023, from https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-281. DOI: 10.1186/1471-2105-15-281
66 Sayyari, E., & Mirarab, S. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Molecular Biology and Evolution, 2016, 33(7): 1654–1668. Retrieved February 28, 2023, from https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw079. DOI: 10.1093/molbev/msw079
67 Schubnel, T., Desutter, L., Garrouste, R., hervé, sophie, & Nel, A. Paleocene of Menat Formation, France, reveals an extraordinary diversity of orthopterans and the last known survivor of a Mesozoic Elcanidae. Acta Palaeontologica Polonica, 2020, 65. Retrieved February 28, 2023, from http://www.app.pan.pl/article/item/app06762019.html. DOI: 10.4202/app.00676.2019
68 Siebert, S., & Backofen, R. MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. Bioinformatics, 2005, 21(16): 3352–3359. Retrieved February 28, 2023, from https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti550. DOI: 10.1093/bioinformatics/bti550
69 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., & Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 2015, 31(19): 3210–3212. Retrieved February 28, 2023, from https://academic.oup.com/bioinformatics/article/31/19/3210/211866. DOI: 10.1093/bioinformatics/btv351
70 Smith, D.R. A synopsis of the sawflies (Hymenoptera: Symphyta) of America south of the United States: introduction, Xyelidae, Pamphiliidae, Cimbicidae, Diprionidae, Xiphydriidae, Siricidae, Orussidae, Cephidae. Systematic Entomology, 1988, 13(2): 205–261. Retrieved December 26, 2021, from https://onlinelibrary.wiley.com/doi/10.1111/j.1365-3113.1988.tb00242.x. DOI: 10.1111/j.1365-3113.1988.tb00242.x
71 Song, L., Florea, L., & Langmead, B. Lighter: fast and memory-efficient sequencing error correction without counting. Genome Biology, 2014, 15(11): 509. Retrieved February 27, 2023, from http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0509-9. DOI: 10.1186/s13059-014-0509-9
72 Song, S.-N., Tang, P., Wei, S.-J., & Chen, X.-X. Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans. Scientific Reports, 2016 a, 6(1): 20972. Retrieved April 19, 2022, from http://www.nature.com/articles/srep20972. DOI: 10.1038/srep20972
73 Song, S.-N., Tang, P., Wei, S.-J., & Chen, X.-X. Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans. Scientific Reports, 2016 b, 6(1): 20972. Retrieved February 28, 2023, from https://www.nature.com/articles/srep20972. DOI: 10.1038/srep20972
74 Song, S.-N., Wang, Z.-H., Li, Y., Wei, S.-J., & Chen, X.-X. The mitochondrial genome of *Tenthredo tienmushana* (Takeuchi) and a related phylogenetic analysis of the sawflies (Insecta: Hymenoptera). Mitochondrial DNA Part A, 2016 a, 27(4): 2860–2861. Retrieved April 19, 2022, from https://www.tandfonline.com/doi/full/10.3109/19401736.2015.1053129. DOI: 10.3109/19401736.2015.1053129
75 Song, S.-N., Wang, Z.-H., Li, Y., Wei, S.-J., & Chen, X.-X. The mitochondrial genome of *Tenthredo tienmushana* (Takeuchi) and a related phylogenetic analysis of the sawflies (Insecta: Hymenoptera). Mitochondrial DNA Part A, 2016 b, 27(4): 2860–2861. Retrieved

chinaXiv:202303.00021v1

February 28, 2023, from https://www.tandfonline.com/doi/full/10.3109/19401736.2015.1053129. DOI: 10.3109/19401736.2015.1053129
76 Suyama, M., Torrents, D., & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research, 2006, 34(Web Server): W609–W612. Retrieved February 28, 2023, from https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkl315. DOI: 10.1093/nar/gkl315
77 Taeger, A. Bestimmungsschlüssel der Keulhornblattwespen Deutschlands (Hymenoptera: Cimbicidae). Kommentierte Bestandsaufnahme. Goecke & Evers, 1998, 193–205.
78 Taeger, A., Blank, S.M., & Liston, A.D. World Catalog of Symphyta (Hymenoptera). Zootaxa, 2010, 2580(1): 1.
79 Tang, P., Zhu, J., Zheng, B., Wei, S., Sharkey, M., Chen, X., & Vogler, A.P. Mitochondrial phylogenomics of the Hymenoptera. Molecular Phylogenetics and Evolution, 2019, 1318–18. Retrieved April 19, 2022, from https://linkinghub.elsevier.com/retrieve/pii/S1055790318303865. DOI: 10.1016/j.ympev.2018.10.040
80 Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B.Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research, 2016, 44(W1): W232–W235. Retrieved February 28, 2023, from https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw256. DOI: 10.1093/nar/gkw256
81 Vaidya, G., Lohman, D.J., & Meier, R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics, 2011, 27(2): 171–180. Retrieved February 28, 2023, from https://onlinelibrary.wiley.com/doi/10.1111/j.1096-0031.2010.00329.x. DOI: 10.1111/j.1096-0031.2010.00329.x
82 Vargas, P., Fernández-Mazuecos, M., & Heleno, R. Phylogenetic evidence for a Miocene origin of Mediterranean lineages: species diversity, reproductive traits and geographical isolation. (J. D. Thompson, Ed.) Plant Biology, 2018, 20157–165. Retrieved February 28, 2023, from https://onlinelibrary.wiley.com/doi/10.1111/plb.12626. DOI: 10.1111/plb.12626
83 Vilhelmsen, L. Giant sawflies and their kin: morphological phylogeny of Cimbicidae (Hymenoptera): Phylogeny of giant sawflies (Cimbicidae). Systematic Entomology, 2019, 44(1): 103–127. Retrieved April 5, 2022, from https://onlinelibrary.wiley.com/doi/10.1111/syen.12314. DOI: 10.1111/syen.12314
84 Vilhelmsen, L., & Chatzaki, V. Fight or fright: potential drivers of body size evolution in Cimbicidae (Insecta, Hymenoptera). Zoomorphology, 2021, 140(3): 341–352. Retrieved February 27, 2023, from https://link.springer.com/10.1007/s00435-021-00533-5. DOI: 10.1007/s00435-021-00533-5
85 Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., & Schatz, M.C. GenomeScope: fast reference-free genome profiling from short reads. (B. Berger, Ed.) Bioinformatics, 2017, 33(14): 2202–2204. Retrieved February 28, 2023, from https://academic.oup.com/bioinformatics/article/33/14/2202/3089939. DOI: 10.1093/bioinformatics/btx153
86 Wan, S., Wei, M., & Niu, G. The complete mitochondrial genome sequence of *Sinopoppia nigroflagella* Wei, 1997 (Hymenoptera: Tenthredinidae) reveals a new gene order. Mitochondrial DNA Part B, 2021, 6(3): 999–1000. Retrieved February 28, 2023, from https://www.tandfonline.com/doi/full/10.1080/23802359.2021.1891989. DOI: 10.1080/23802359.2021.1891989

chinaXiv:202303.00021v1

87 Wei M., Nei H., & Niu G. Sympatricspeciation：aprincipalpatternofspeciation？. Journal of Central South University of Forestry& Technology, 2010, 30(3): 1–11. Retrieved from https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2010&filename=ZNLB201003003&v=. DOI: 10.14067/j.cnki.1673-923x.2010.03.002
88 Wei, M., Wu, X., Niu, G., & Xin, H. Rediscovery of Odontocimbex svenhedini Malaise (Hymenoptera: Cimbicidae) from China with description of the female and a key to Asian genera of Cimbicidae. Entomotaxonomia, 2012, 34(2): 435–441.
89 Wu, D., Wang, H., Wei, M., & Niu, G. The nearly complete mitochondrial genome of *Colochela zhongi* Wei, 2016 (Hymenoptera: Tenthredinidae) and phylogenetic analysis. Mitochondrial DNA Part B, 2020, 5(3): 3341–3342. Retrieved February 28, 2023, from https://www.tandfonline.com/doi/full/10.1080/23802359.2020.1820393. DOI: 10.1080/23802359.2020.1820393
90 Wu, R., Wei, M., Liu, M., & Niu, G. Advancement in sequencing the mitochondrial genome of *Birmella discoidalisa* Wei, 1994 (Hymenoptera: Tenthredinidae) and the phylogenetic classification of Fenusini. Mitochondrial DNA Part B, 2019, 4(2): 4100–4101. Retrieved February 28, 2023, from https://www.tandfonline.com/doi/full/10.1080/23802359.2019.1692728. DOI: 10.1080/23802359.2019.1692728
91 Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. An index of substitution saturation and its application. Molecular Phylogenetics and Evolution, 2003, 26(1): 1–7. Retrieved February 28, 2023, from https://linkinghub.elsevier.com/retrieve/pii/S1055790302003263. DOI: 10.1016/S1055-7903(02)00326-3
92 Yan Y. *Phylogeny and Taxonomy of Cimbicidae (Hymenoptera: Tenthredinoidea)* (PhD's Thesis). Central South University of Forestry and Technology. Retrieved from https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2022&filename=1021128354.nh&v=
93 Yan Y., Li Z., Li G., & Wei M. A New Species of Zaraea(Leach，1817) (Hymenoptera，Cimbicidae) from China Harming Padus brachypoda. Scientia Silvae Sinicae, 2020, 56(5): 105–112. Retrieved from https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2020&filename=LYKE202005012&v=.
94 Yan, Y., Niu, G., Zhang, Y., Ren, Q., Du, S., Lan, B., & Wei, M. Complete mitochondrial genome sequence of *Labriocimbex sinicus* , a new genus and new species of Cimbicidae (Hymenoptera) from China. PeerJ, 2019, 7e7853. Retrieved December 8, 2021, from https://peerj.com/articles/7853. DOI: 10.7717/peerj.7853
95 Yang, C., Shan, B., Liu, Y., Wang, L., Wu, Q., Luo, Z., & Sun, D. Complete Mitochondrial Genome of Two Ectoparasitic Capsalids (Platyhelminthes: Monogenea: Monopisthocotylea): Gene Content, Composition, and Rearrangement. Genes, 2022, 13(8): 1376. Retrieved August 28, 2022, from https://www.mdpi.com/2073-4425/13/8/1376. DOI: 10.3390/genes13081376
96 Yang, J., Sun, Z., Wei, M., & Niu, G. The complete mitochondrial genome of *Allantus togatus* (Panzer, 1801), in view of possible cryptic species. Mitochondrial DNA Part B, 2021, 6(3): 1114–1115. Retrieved December 8, 2021, from https://www.tandfonline.com/doi/full/10.1080/23802359.2021.1899875. DOI: 10.1080/23802359.2021.1899875
97 Zhang, H.-L., Wang, R.-M., Han, Z.-P., Zhang, Y.-F., & Liu, J.-X. The first mitochondrial genome for Brahmin moths (Brahmaeidae) and implications for the higher phylogeny of Bombycoidea. Journal of Asia-Pacific Entomology, 2018, 21(2): 578–584. Retrieved November 22, 2021, from

chinaXiv:202303.00021v1

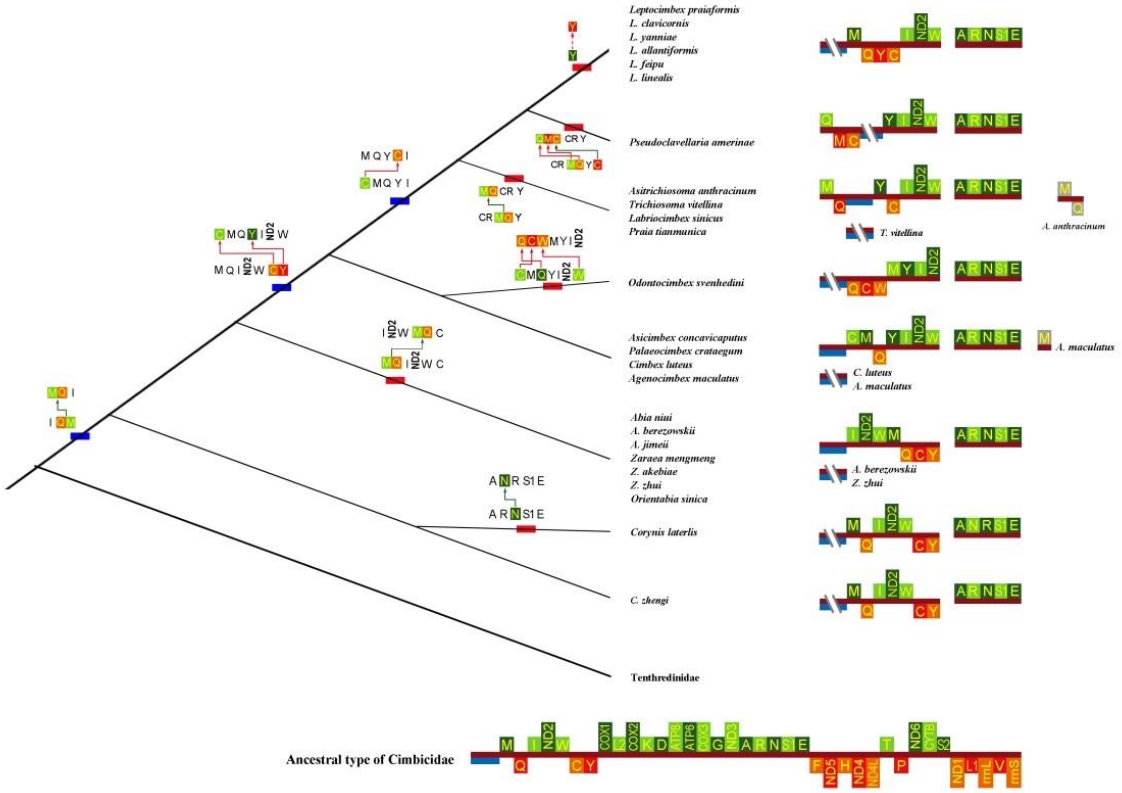
https://linkinghub.elsevier.com/retrieve/pii/S1226861517306337. DOI: 10.1016/j.aspen.2018.03.013

98 Zhang, M., Ruan, Y., Wan, X., Tong, Y., Yang, X., & Bai, M. Geometric morphometric analysis of the pronotum and elytron in stag beetles: insight into its diversity and evolution. ZooKeys, 2019, 83321–40. Retrieved November 22, 2021, from https://zookeys.pensoft.net/article/26164/. DOI: 10.3897/zookeys.833.26164

chinaXiv:202303.00021v1

Figuers

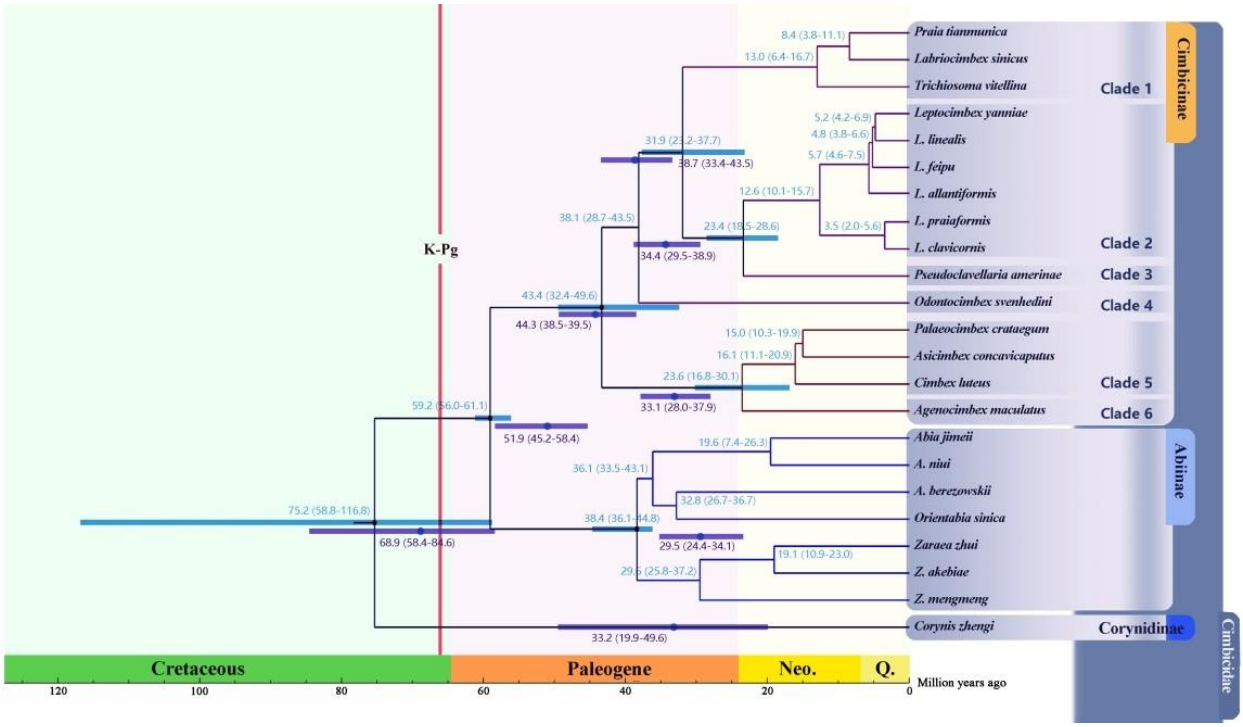
fig. 1. Genetic rearrangement tree of the Cimbicidae. Genes that are transcribed from the J- and N-strands are shown in green and orange, respectively. The tRNA genes are labeled by their single-letter amino acid code. The arrows indicate gene rearrangement events.



Genetic rearrangement tree of the Cimbicidae. Genes that are transcribed from the J- and N-strands are shown in green and orange, respectively. The tRNA genes are labeled by their single-letter amino acid code. The arrows indicate gene rearrangement events.

chinaXiv:202303.00021v1

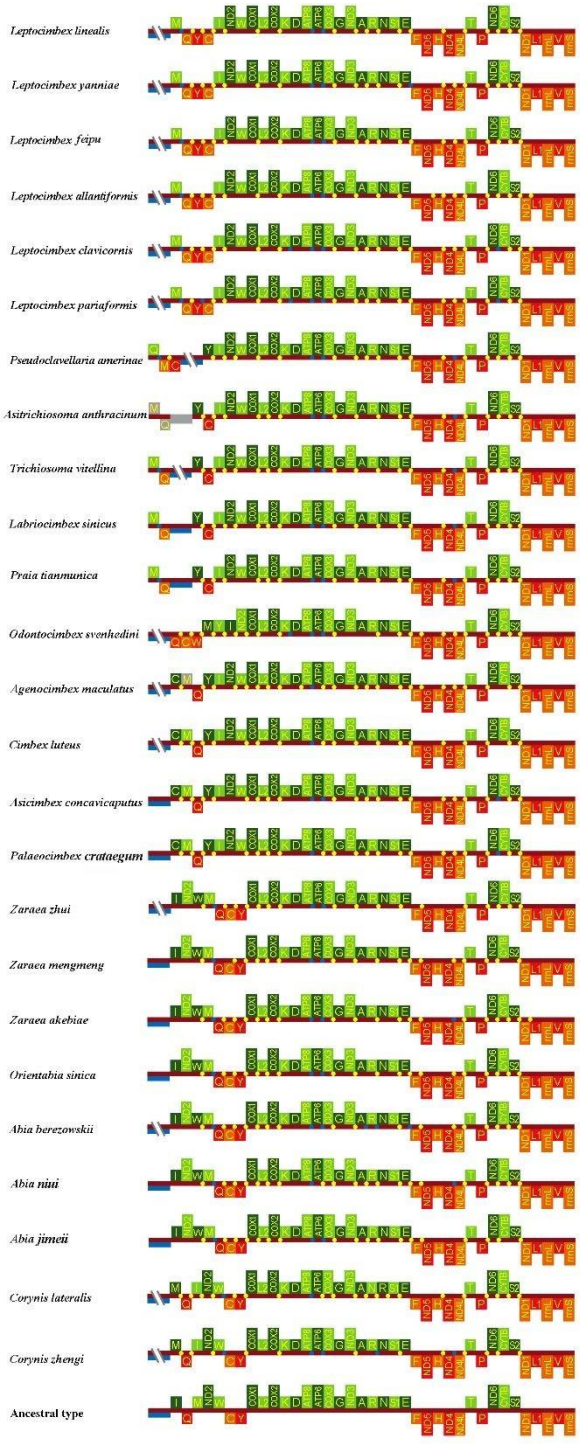
Fig. 2. Chronogram of the Cimbicidae resulting from the MCMCtree analysis. Red circles at nodes mark calibration points. Blue bars indicate 95% high posterior density intervals of the age estimates based on the SCOs, and purple bars indicate that based on the mtG. Mean ages of some nodes are shown above the bars. Neo, Neogene; Q,Quaternary.



Chronogram of the Cimbicidae resulting from the MCMCtree analysis. Red circles at nodes mark calibration points. Blue bars indicate 95% high posterior density intervals of the age estimates based on the SCOs, and purple bars indicate that based on the mtG. Mean ages of some nodes are shown above the bars. Neo, Neogene; Q, Quaternary.

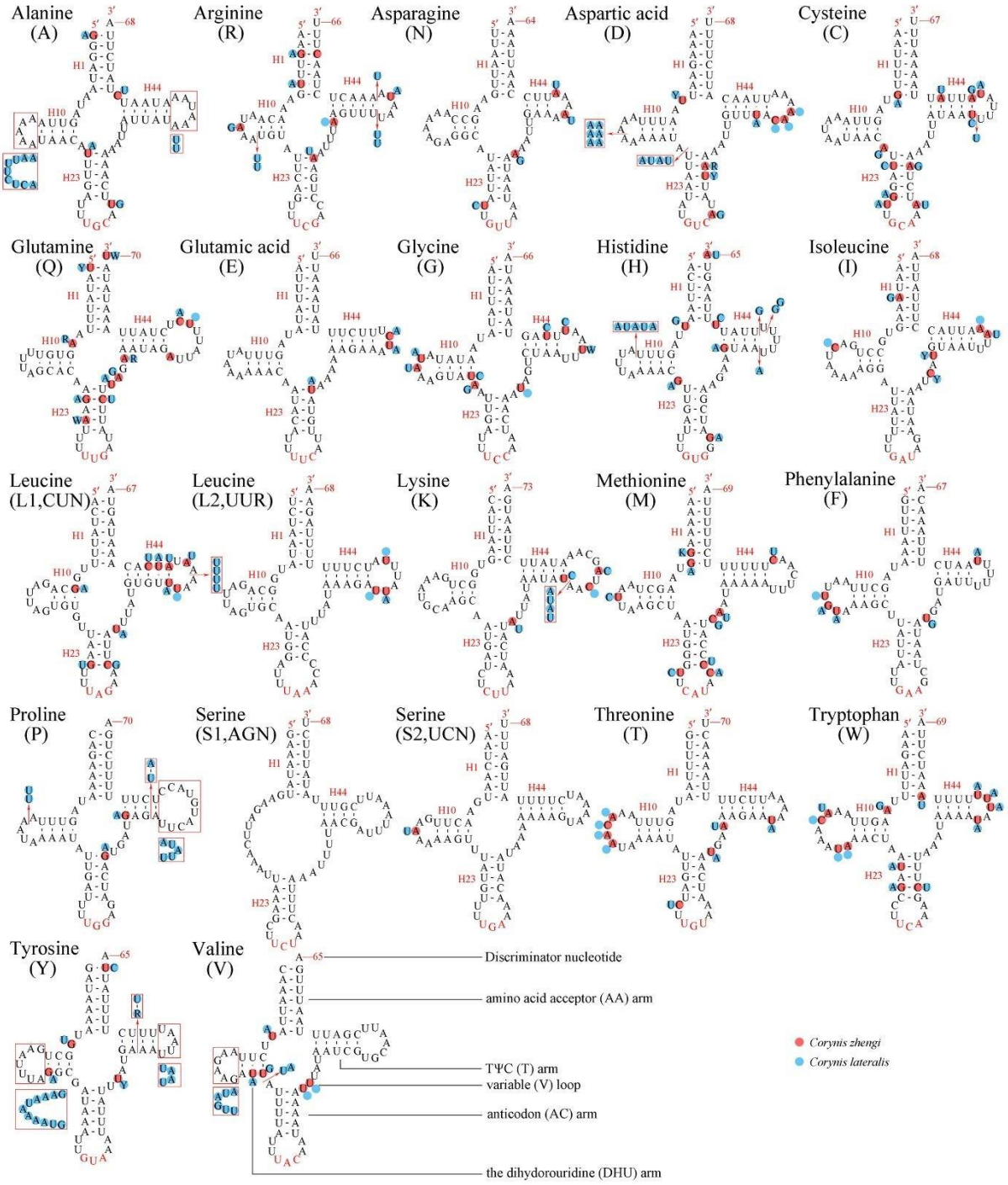
chinaXiv:202303.00021v1

Fig. S1.



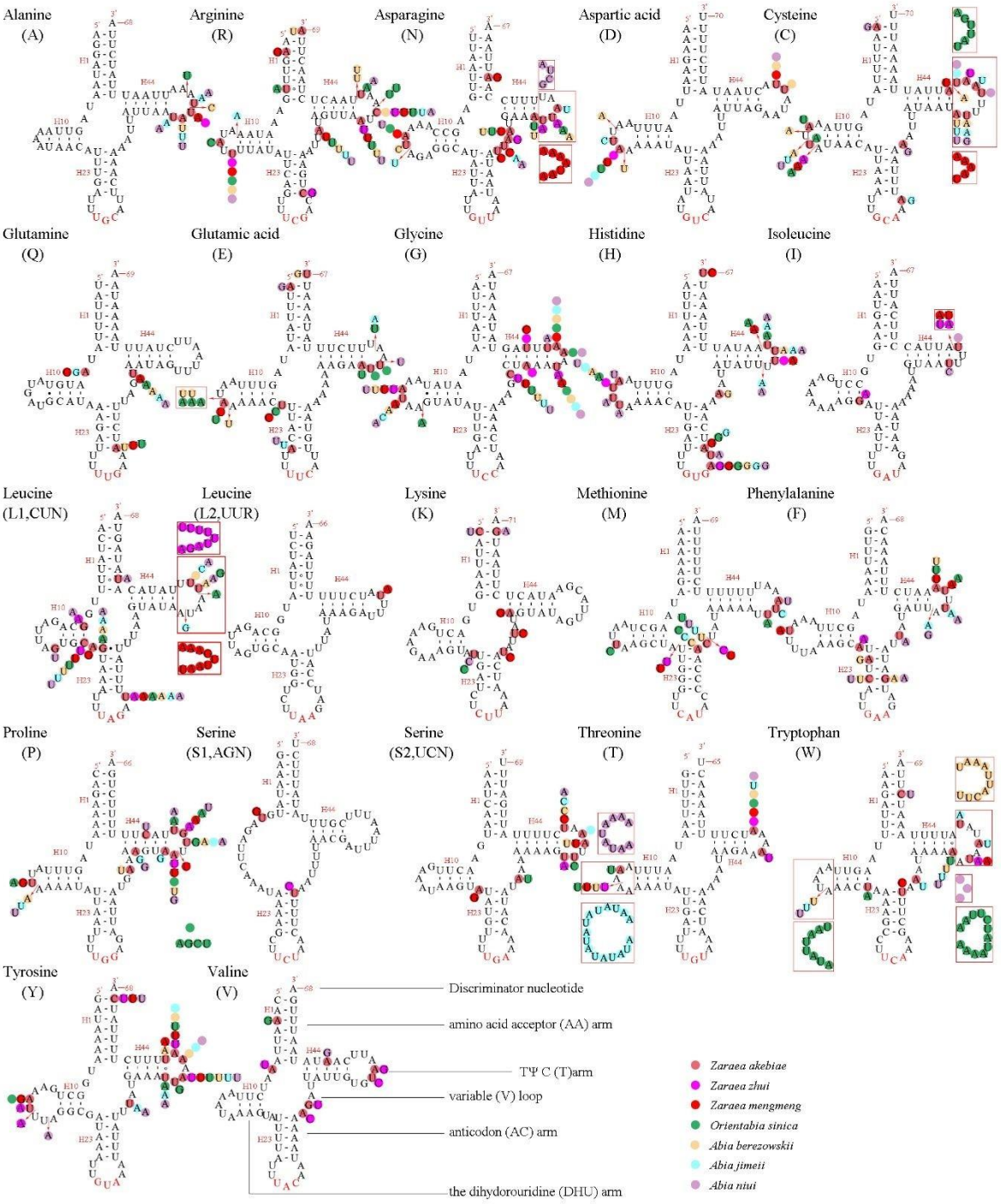
chinaXiv:202303.00021v1

Fig. S2. Secondary structure of tRNA families in mtDNAs of *Corynis*.



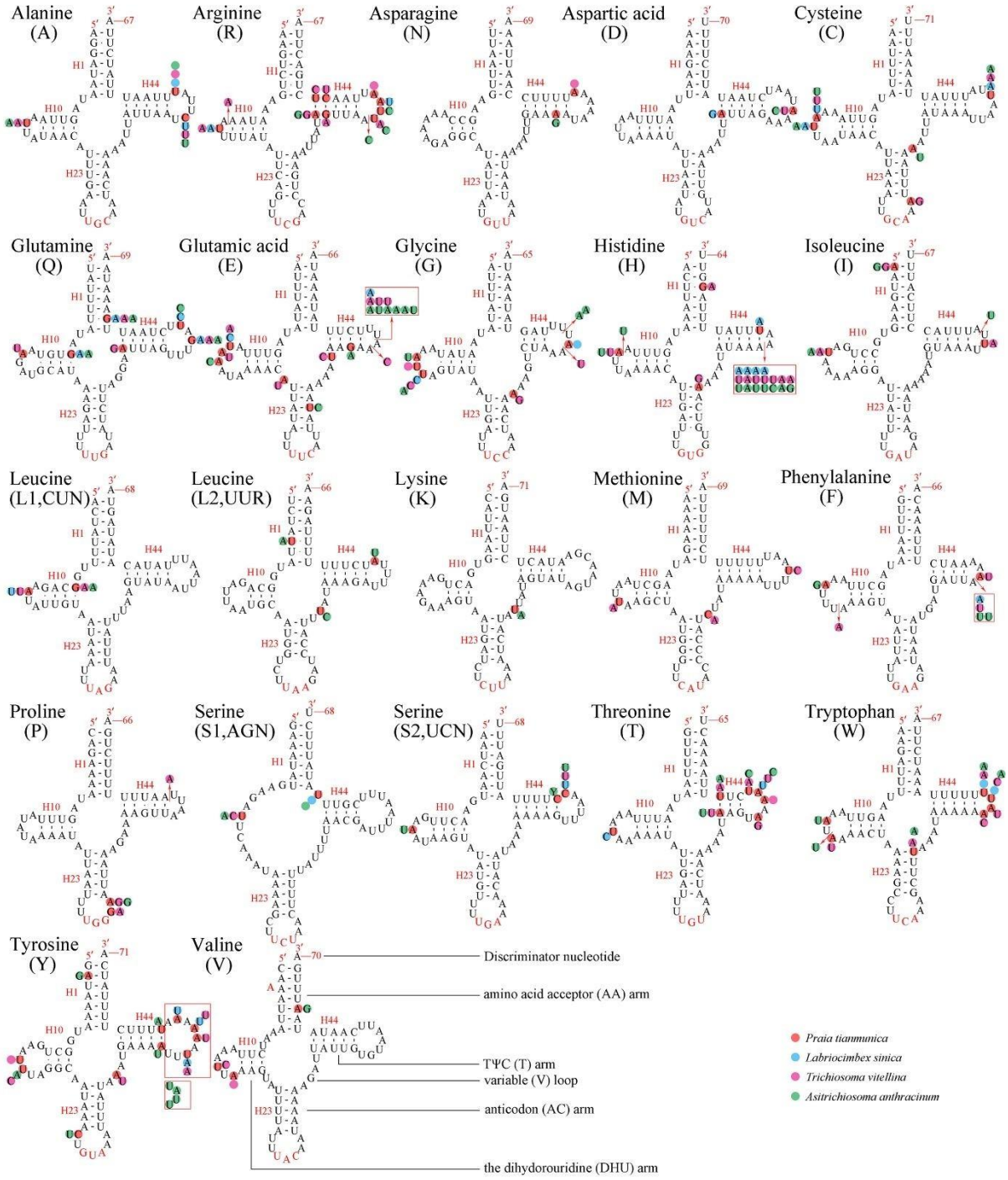
chinaXiv:202303.00021v1

Fig. S3. Secondary structure of tRNA families in mtDNAs of the Abiinae.



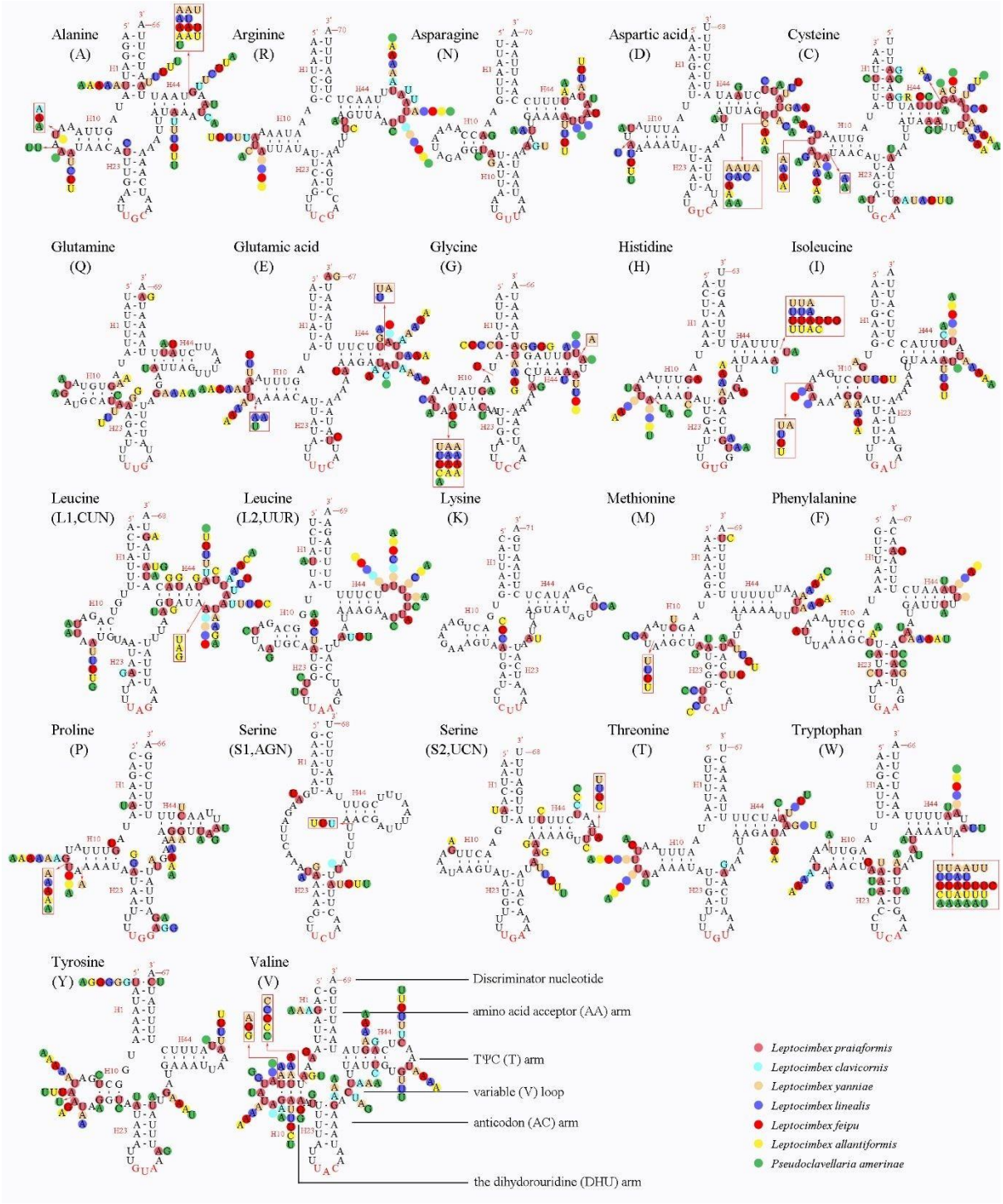
chinaXiv:202303.00021v1

Fig. S4. Secondary structure of tRNA families in mtDNAs of Clade 1 in the Cimbicinae.



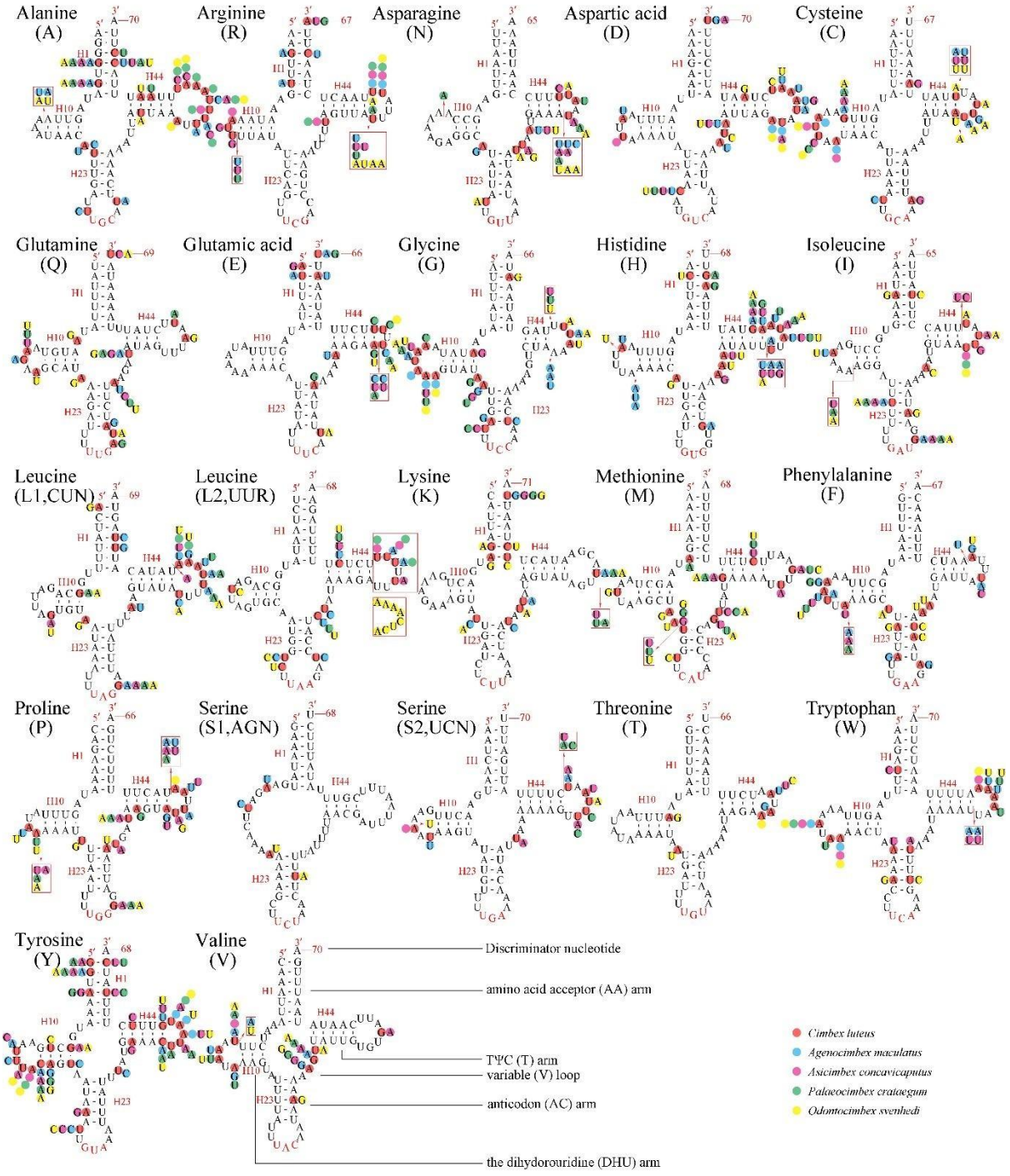
chinaXiv:202303.00021v1

Fig. S5. Secondary structure of tRNA families in mtDNAs of Clade 2 and 3 in Cimbicinae.



chinaXiv:202303.00021v1

Fig. S6. Secondary structure of tRNA families in mtDNAs of Clade 4, 5 and 6 in the Cimbicinae.

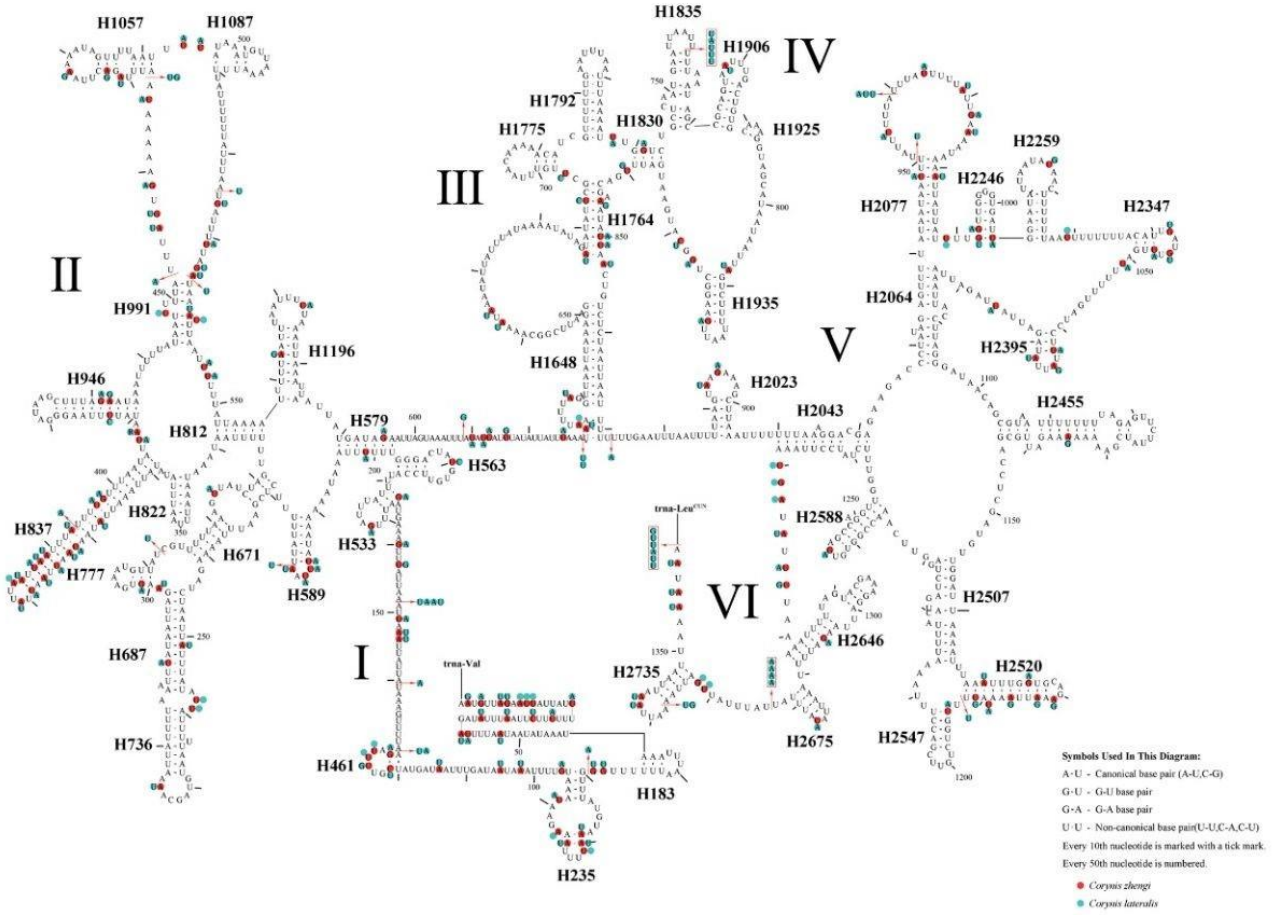


chinaXiv:202303.00021v1

chinaXiv:202303.00021v1

The predicted secondary structures of

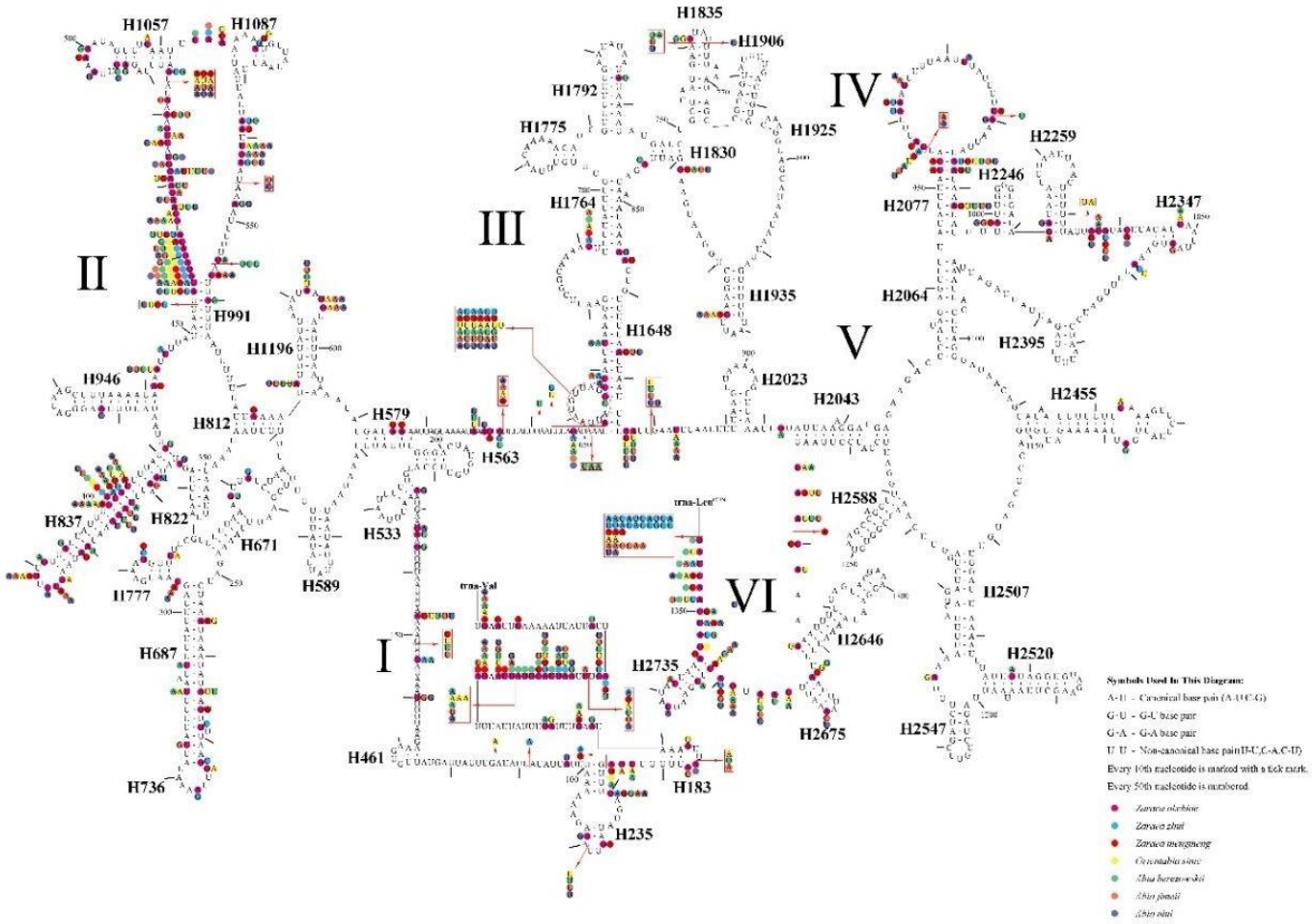
Fig. S7. The predicted secondary structures of *rrnL* of *Corynis*.



chinaXiv:202303.00021v1

The predicted secondary structures of

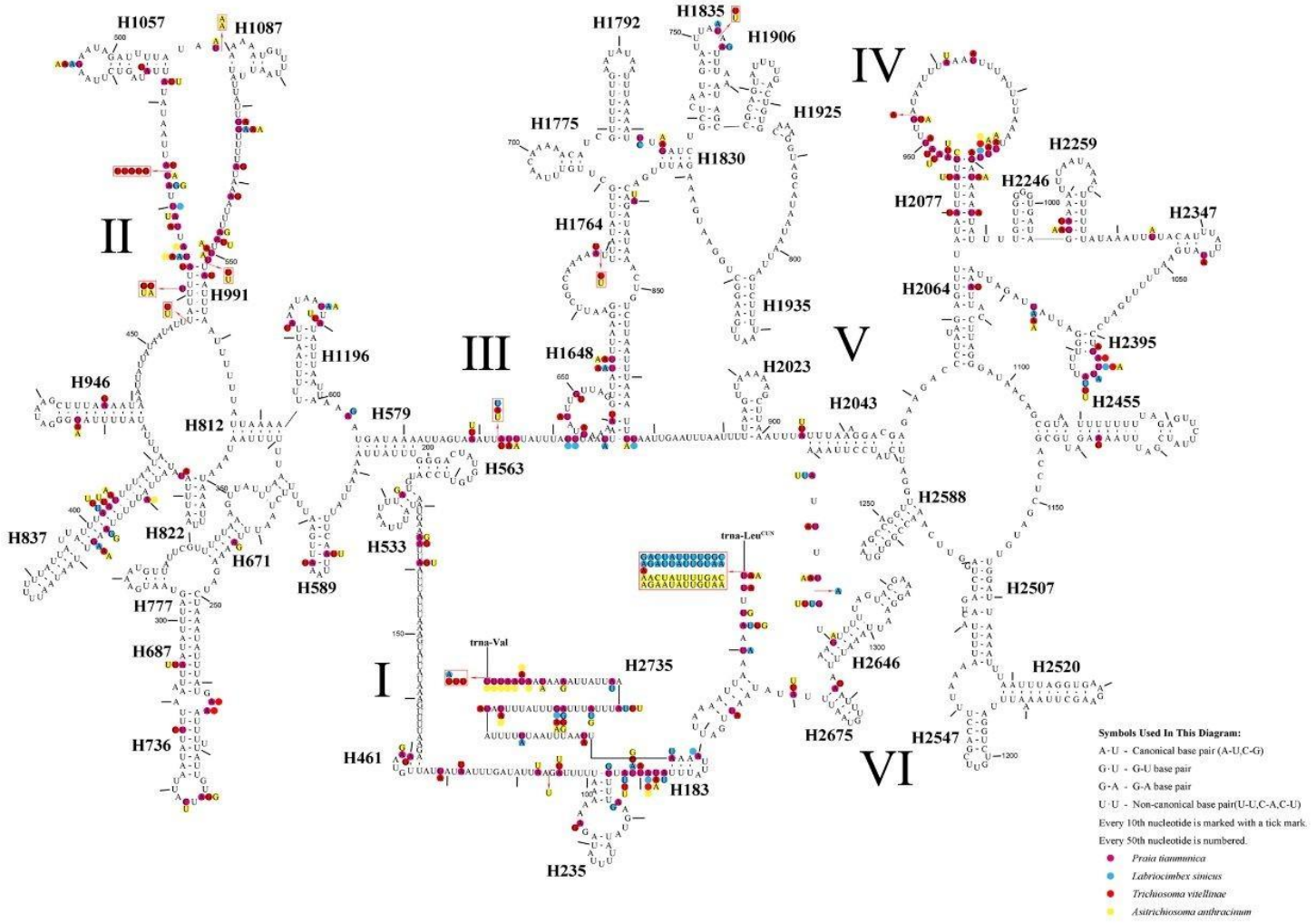
Fig. S8. The predicted secondary structures of *rrnL* of Abiinae



chinaXiv:202303.00021v1

The predicted secondary structures of

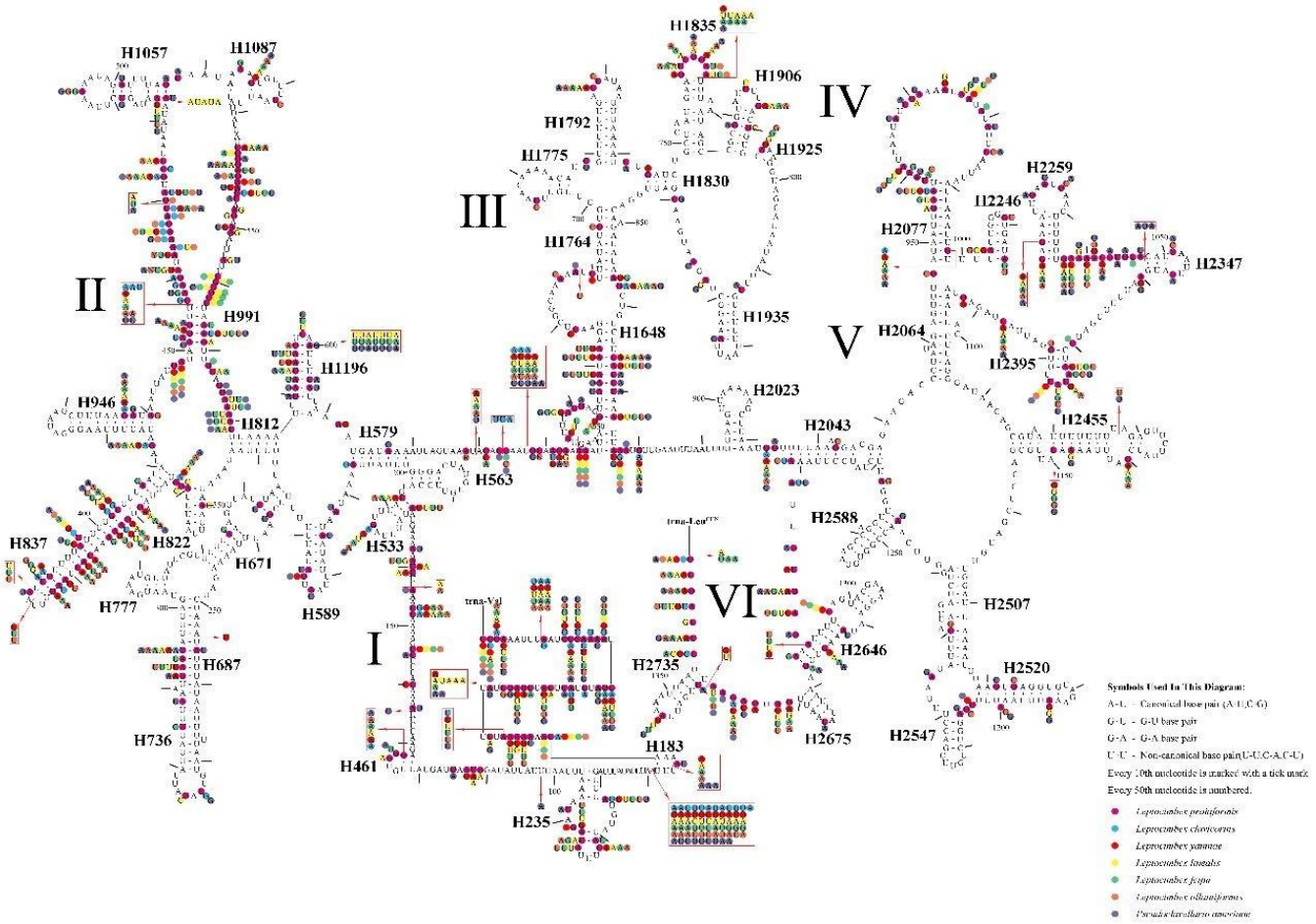
Fig. S9. The predicted secondary structures of *rrnL* of Clade 1 in the Cimbicinae..



chinaXiv:202303.00021v1

The predicted secondary structures of

Fig. S10. The predicted secondary structures of *rrnL* of Clade 2 and 3 in Cimbicinae



chinaXiv:202303.00021v1

The predicted secondary structures of

Fig. S11. The predicted secondary structures of *rrnL* of Clade 4, 5 and 6 in the Cimbicinae.

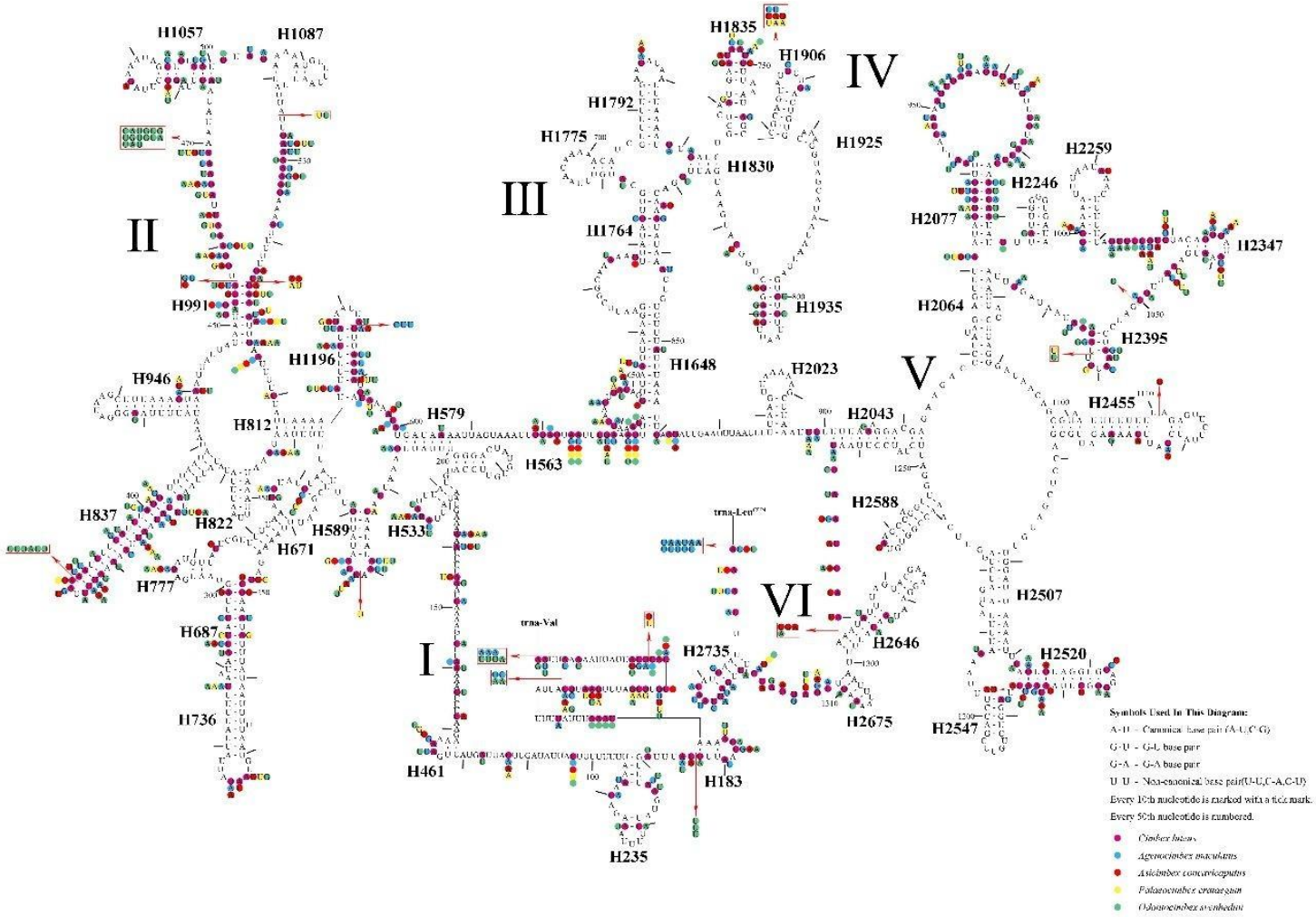
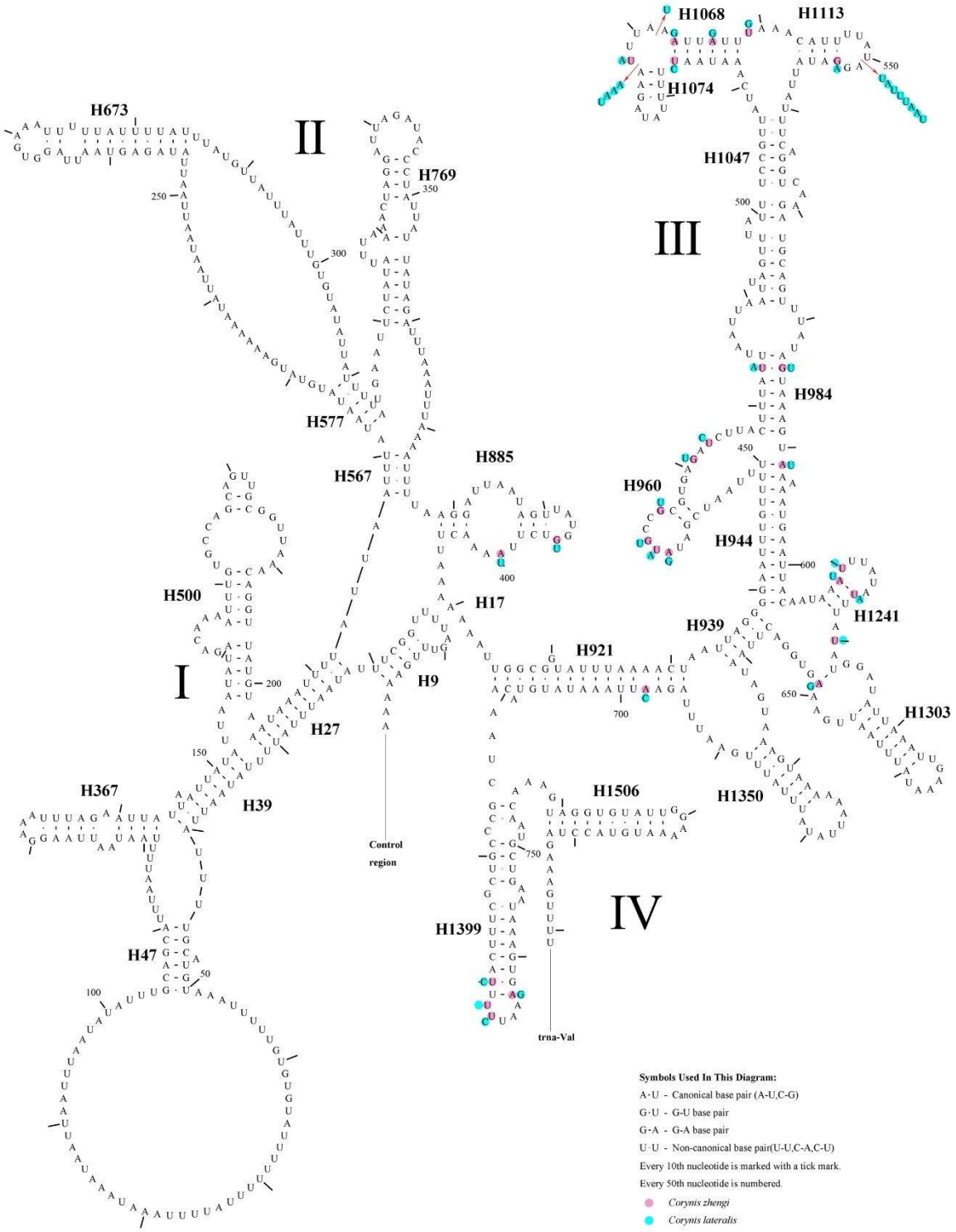
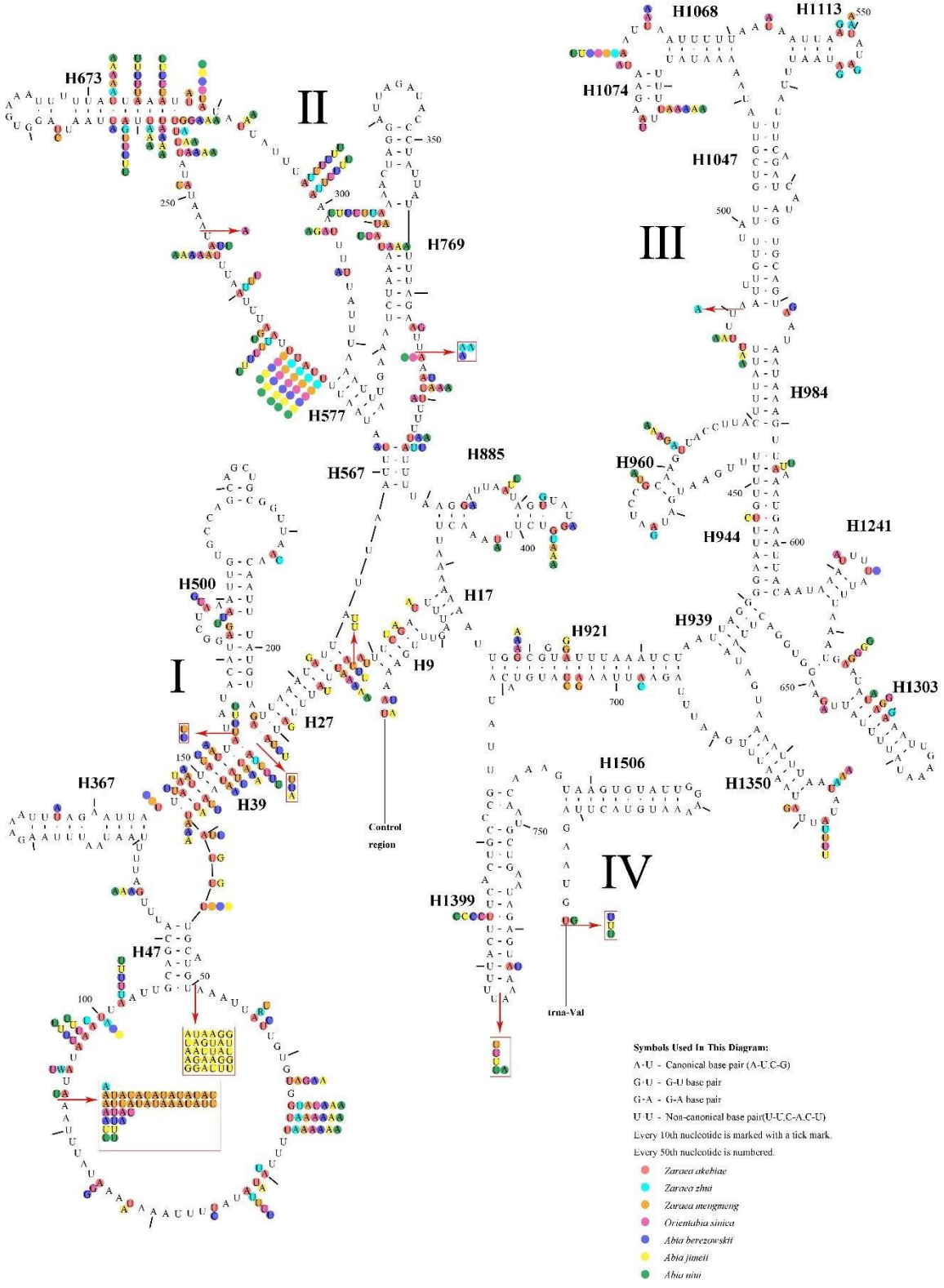


Fig. S12. The predicted secondary structures of *rrnS* of *Corynis*..



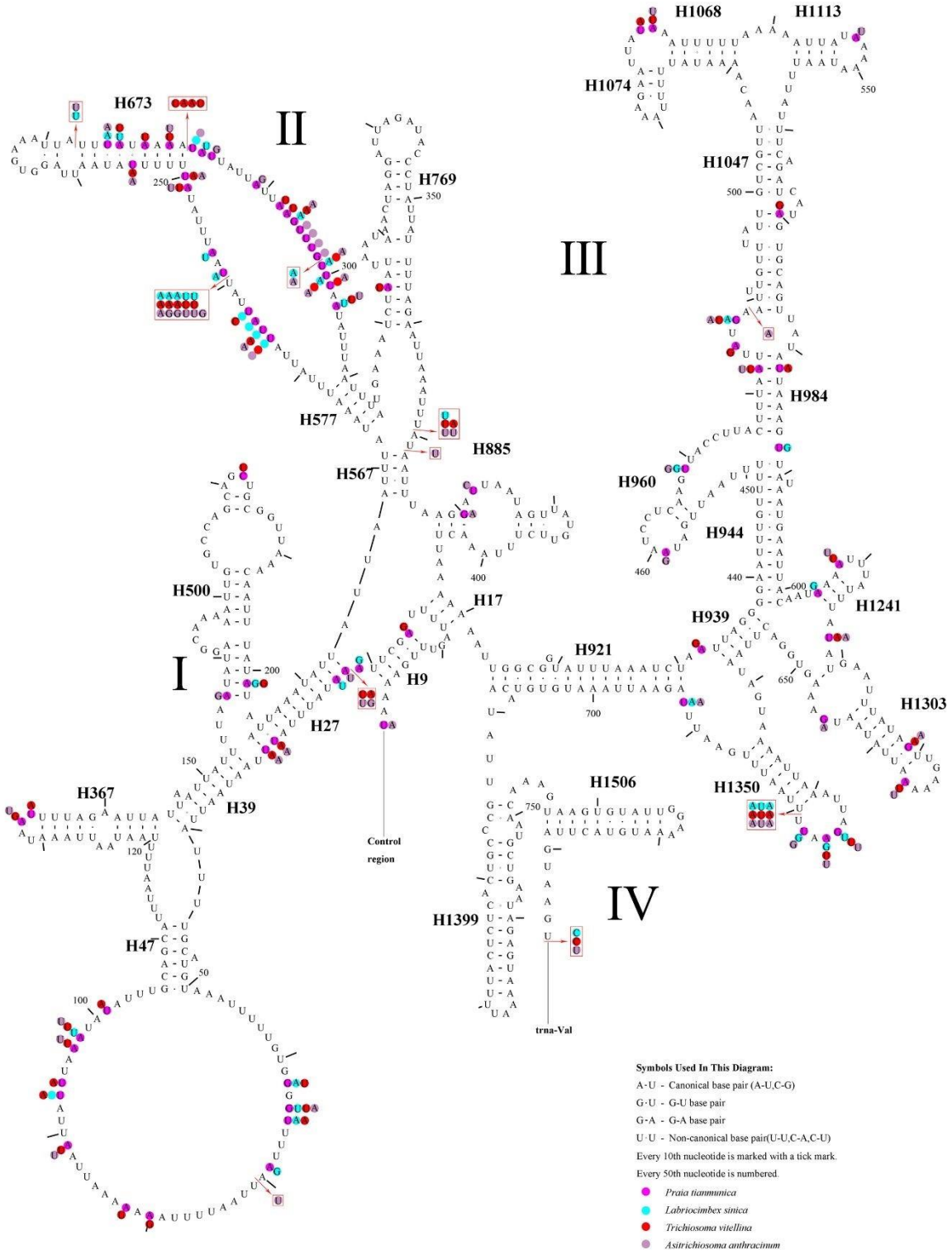
chinaXiv:202303.00021v1

Fig. S13. The predicted secondary structures of *rrnS* of the Abiinae..



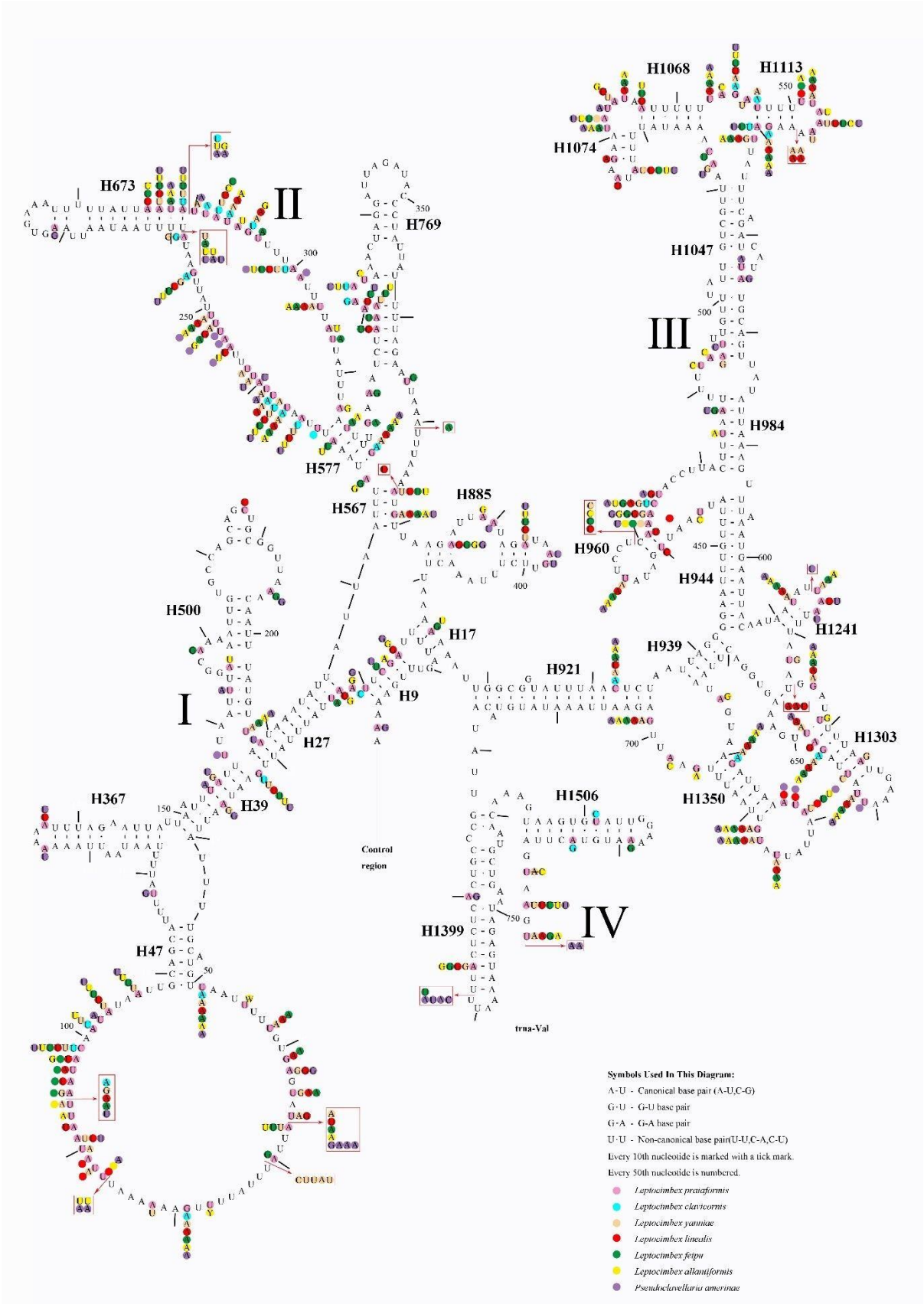
chinaXiv:202303.00021v1

Fig. S14. The predicted secondary structures of *rrnS* of Clade 1 in the Cimbicinae



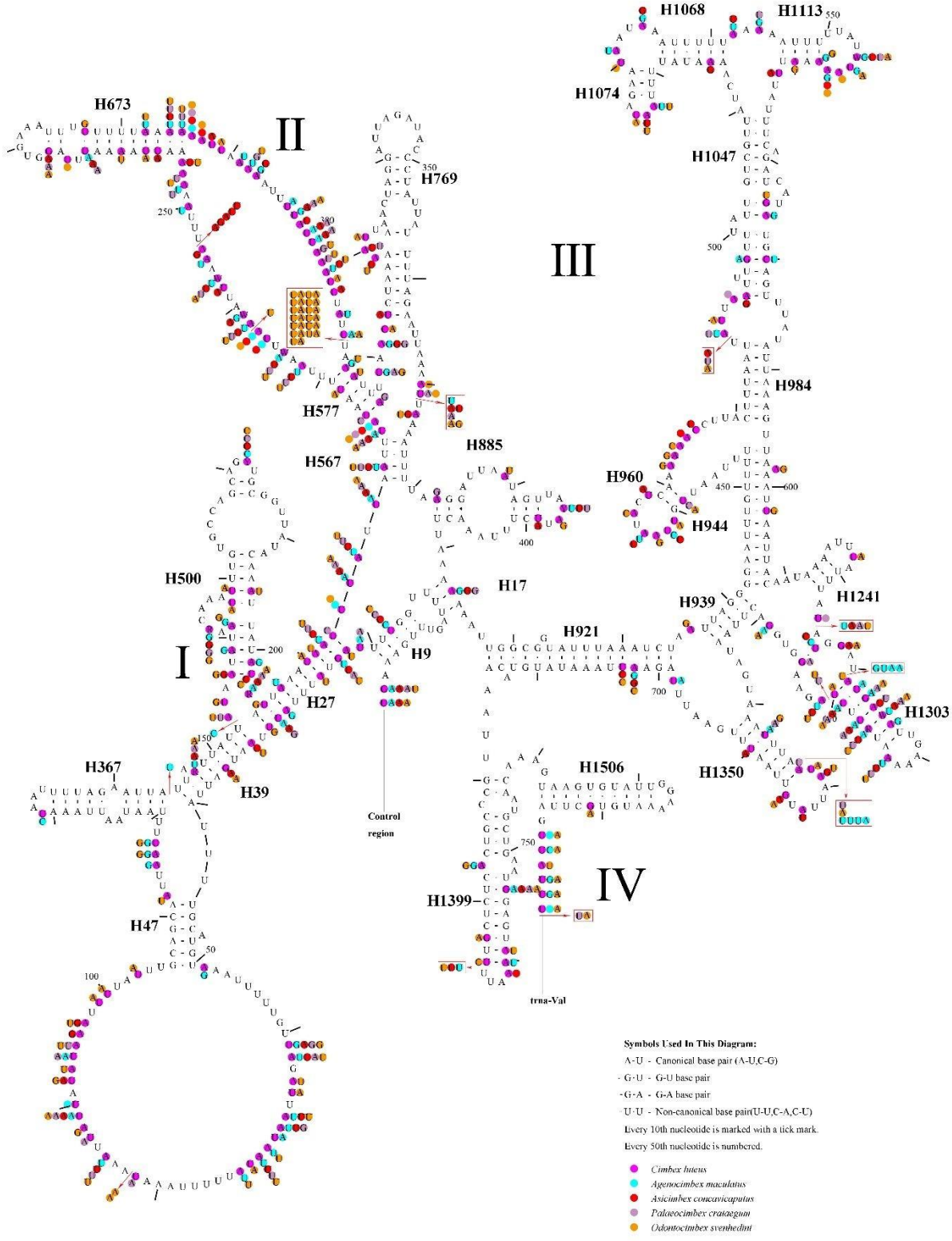
chinaXiv:202303.00021v1

Fig. S15. The predicted secondary structures of *rrnS* of Clade 2 and 3 in Cimbicinae.



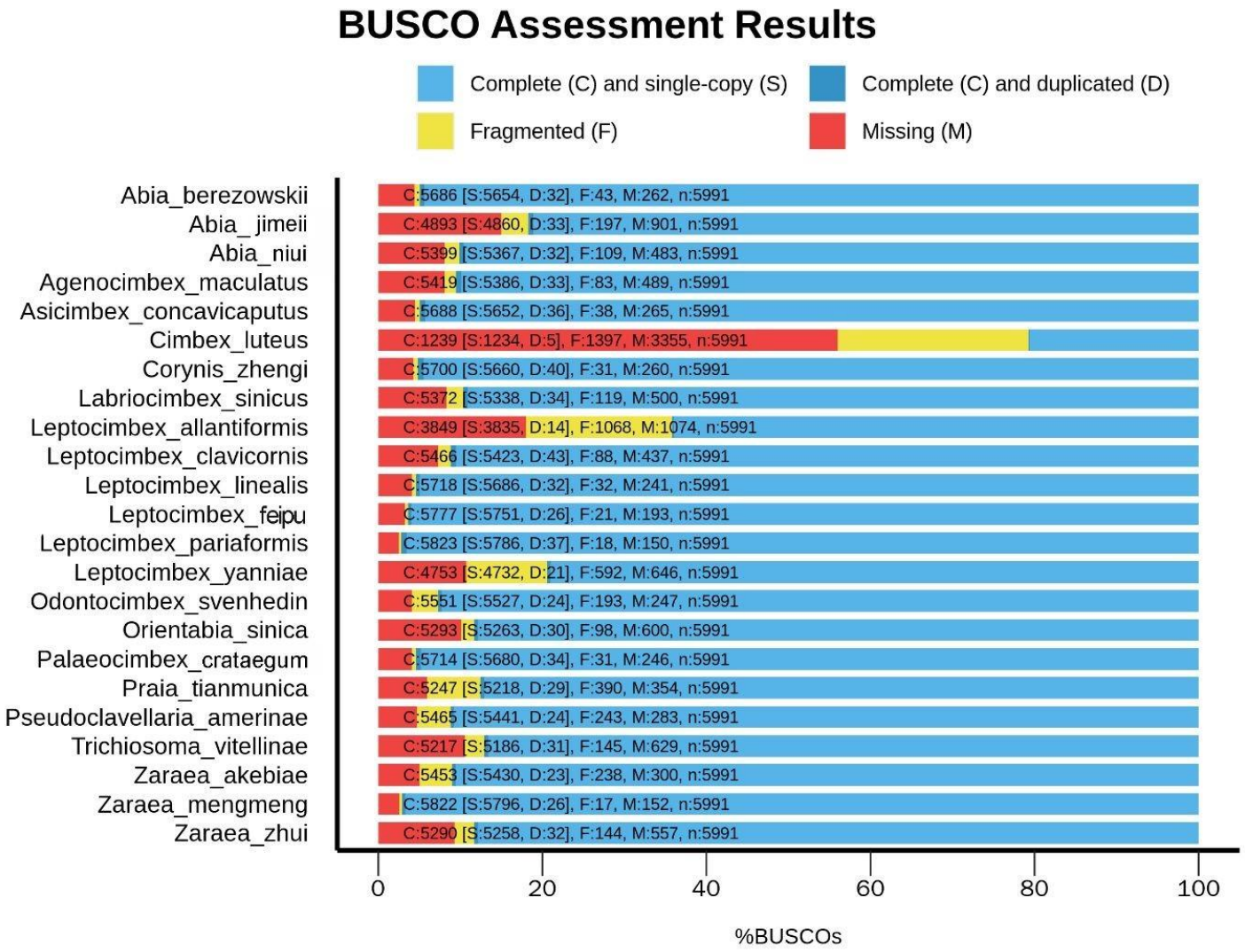
chinaXiv:202303.00021v1

Fig. S16. The predicted secondary structures of rrnS of Clade 4, 5 and 6 in the Cimbicinae..



chinaXiv:202303.00021v1

Fig. S17. BUSCO assessment of Cimbicid genome assemblies.



chinaXiv:202303.00021v1

chinaXiv:202303.00021v1

Fig. S18. Phylogenetic tree of Cimbicidae based on the sequences of 12 unsaturated protein-coding genes in the mitochondrial genome (mtG). Both ML and BI analyses produced the same tree topology. The numbers at the branches represent Maximum-likelihood bootstrap values/Bayesian posterior probabilities. 100/1.00 is denoted by an asterisk (*=1/100). The scale bar indicates the number of substitutions per site.

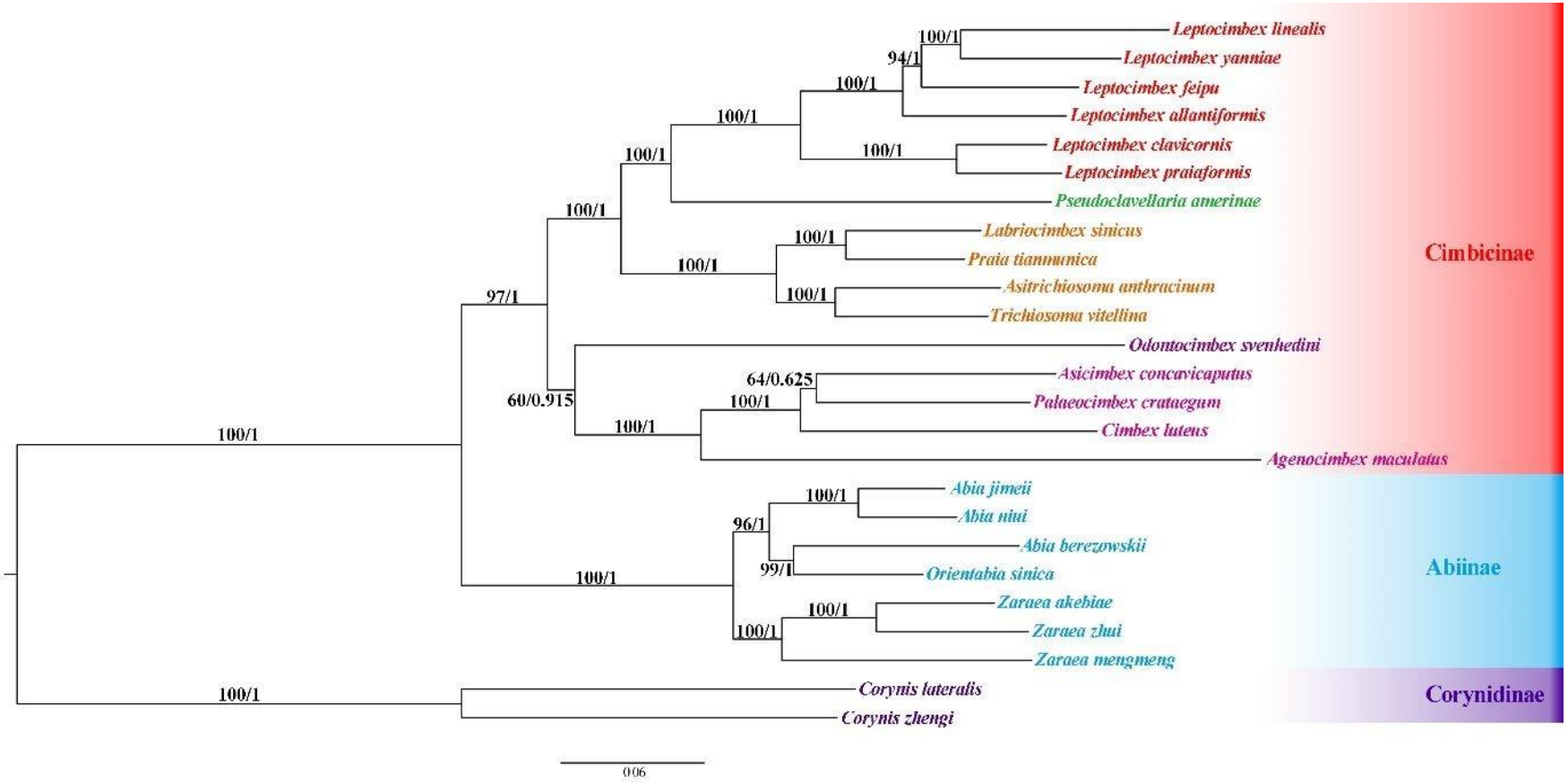
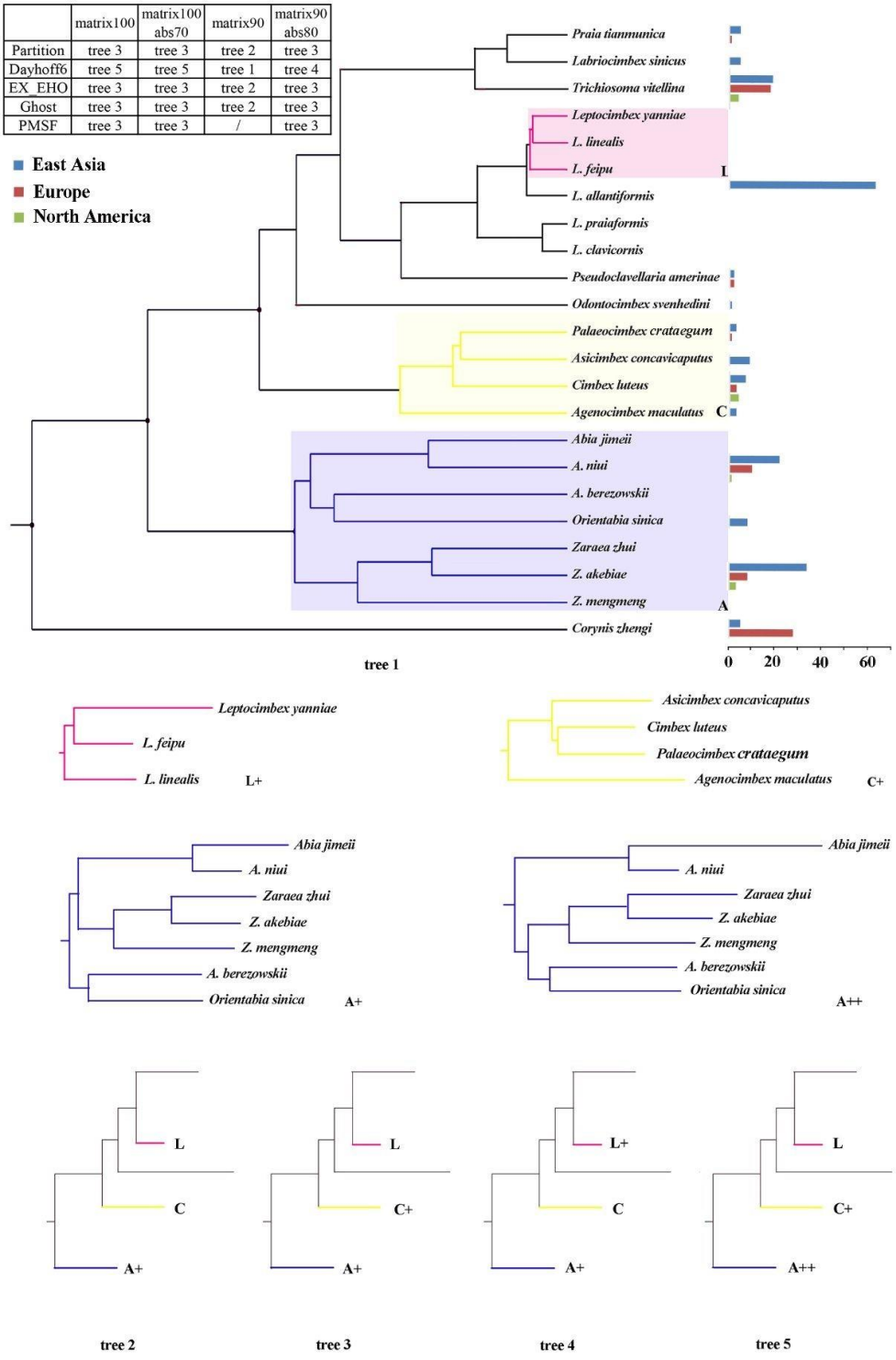


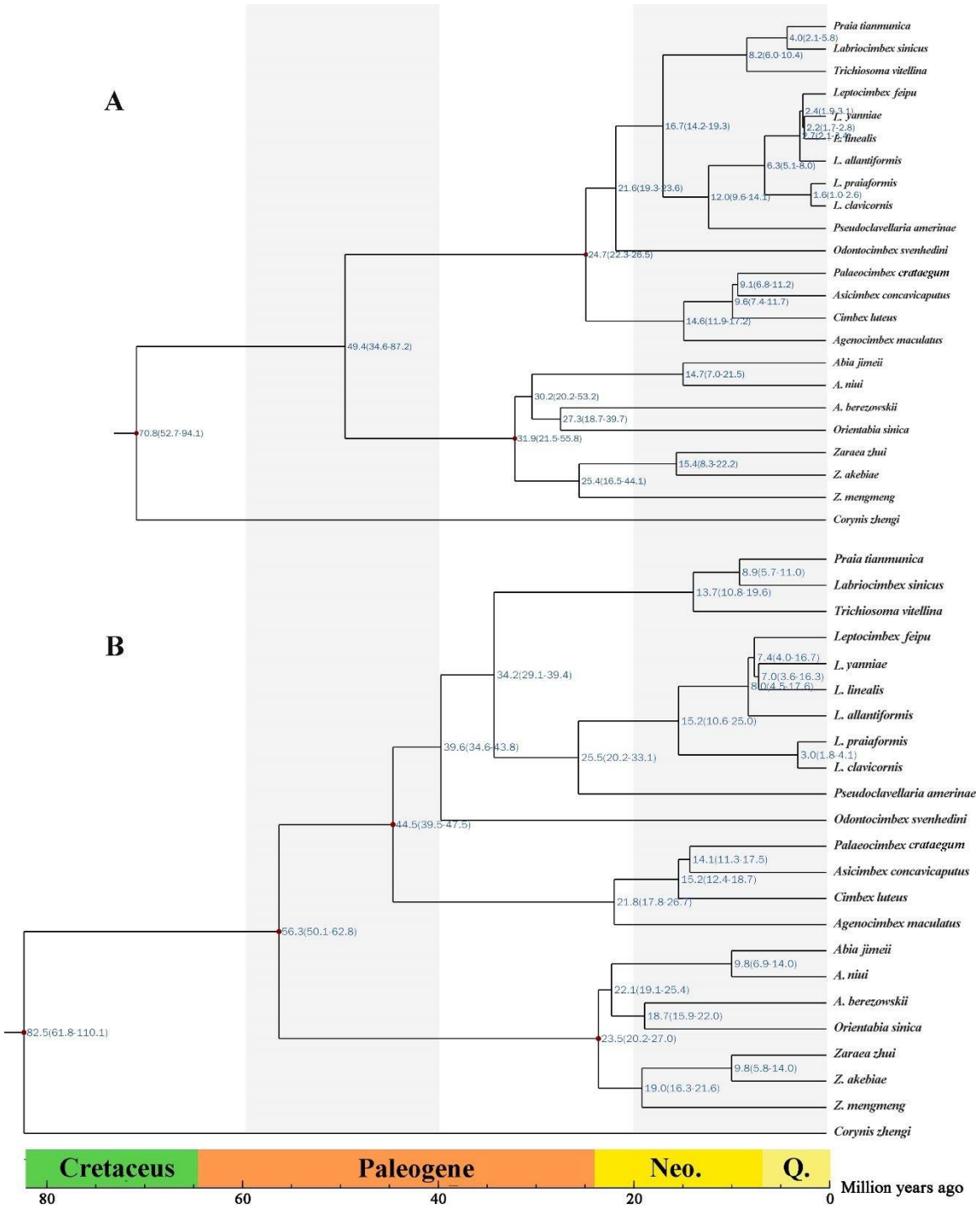
Fig. S19. Summary of Phylogenetic tree topologies of nuclear single-copy orthologs (SCOs) using varies strategy cross all datasets, and a comparison of the taxonomic richness of Cimbicid in East Asia, Europe and North America.

	matrix100	matrix100 abs70	matrix90	matrix90 abs80
Partition	tree 3	tree 3	tree 2	tree 3
Dayhoff6	tree 5	tree 5	tree 1	tree 4
EX_EHO	tree 3	tree 3	tree 2	tree 3
Ghost	tree 3	tree 3	tree 2	tree 3
PMSF	tree 3	tree 3	/	tree 3



chinaXiv:202303.00021v1

Fig. S1. Dated phylogeny constructed with nuclear single-copy orthologs (SCOs) in MCMCtree. The red spots indicate the nodes calibrated by fossils. A was under scheme 1 and B was under scheme 3. The axis in the middle refers to million years and shows the geological time.



chinaXiv:202303.00021v1

Tables

Table S1. Comparison of Cimbicidae divergence time estimates from previous studies

Study	Molecular dataset		Age of Cimbicidae		Age of clade (crown group)			
	No. of taxa (Cimbicidae)	No. of loci	Stem	Crown	Abiinae+Cimbicinae	Abiinae	Cimbicinae	Corynidinae
Ronquist et al., 2012	3	7 (>5000 bp)	200	135		\	\	\
O'Reilly et al., 2015	3	/	160	110		\	\	\
Nyman et al., 2019	8	9 loci	88.26	75.86	49	20	35	0
Niu et al., 2021	9	P123RNA	110.71	91.96		\	57.72	\

chinaXiv:202303.00021v1

Table S2. List of fossils

Geologic ages	Subfamily	Species	Time interval (Ma)	Locality
Miocene	Cimbicinae	*Cimbex turoliana* Riou, 1992	8.7–5.333	Ardèche, France
		Cimbex miocenica Riou, 1992	8.7–5.333	Ardèche, France
		Cimbex chromoptera Zhang, 1989	20.44–15.97	Shanwang, Shandong, China
		Clavellaria bicolor Zhang, 1989	20.44–15.97	Shanwang, Shandong, China
		Clavellaria longiclava Zhang, 1989	20.44–15.97	Shanwang, Shandong, China
		Clavellaria shandogensis (Hong et Wang, 1985)	20.44–15.97	Shanwang, Shandong, China
		Clavellaria molpa Zhang, Sun & Zhang, 1994	20.44–15.97	Shanwang, Shandong, China
		Sinocimbex silacea Zhang, Sun & Zhang, 1994	20.44–15.97	Shanwang, Shandong, China
		Sinocimbex pellucida Zhang, Sun & Zhang, 1994	20.44–15.97	Shanwang, Shandong, China
		Cimbex sp. Fujiyama, 1985	23.03–15.97	Seki, Sado Island, Japan
		Trichiosomites obliviosus Brues, 1908	37.2–33.9	Florissant, Colorado, USA
		Pseudocimbex clavatus Rohwer, 1908	37.2–33.9	Florissant, Colorado, USA
		Cimbex vetusculus Cockerell, 1922	37.2–33.9	Florissant, Colorado, USA
		Phenacoperga coloradensis Cockerell, 1908	/	/
		14 samples belongs to Cimbicinae, Coryninae or Pachylostictinae	47.8-56	Okanagan Highland, Canada
		Eopachylosticta byrami (Cockerell, 1925)	50.3–46.2	Green River, Colorado, USA
Paleocene		*Cenocimbex menatensis* Nel, 2004	61.6–59.2	Menat, Auvergne, France
	Abiinae	*Zaraea* sp.	3.6–2.588	Niedersachsen, Germany
		Abia shandongensis (Hong et Wang, 1985)	20.44–15.97	Shanwang, Shandong, China
		Abia paurocephala Zhang, 1989	20.44–15.97	Shanwang, Shandong, China
		Abia maculosa Zhang, 1989	20.44–15.97	Shanwang, Shandong, China
		Abia cf lonicerae Zhang, 1994	20.44–15.97	Shanwang, Shandong, China
		Abia shanwangensis (Hong, 1984)	20.44–15.97	Shanwang, Shandong, China
	Cimbicidae/Diprionidae	*incertae sedis*	56-47.8	Okanagan Highland, Canada

chinaXiv:202303.00021v1

Table S3. List of taxa used in analyses, indicating data type (mt genome) and origin (current study or publicly available data)

brary ID	subfamily	Species	GenBank number	Reference
CSCS-Hym-MC0047	Abiinae	*Abia berezowskii*	OM066090	Current study
CSCS-Hym-MC0164		*Abia jimeii*	OM066092	Current study
CSCS-Hym-MC0169		*Abia niui*	OL549452	Current study
CSCS-Hym-MC0043		*Orientabia sinica*	OM066089	Current study
CSCS-Hym-MC0131		*Zaraea akebiae*	OM066097	Current study
CSCS-Hym-MC0046		*Zaraea mengmeng*	OM066095	Current study
CSCS-Hym-MC0163		*Zaraea zhui*	OL549453	Current study
CSCS-Hym-MC0048	Cimbicidae	*Agenocimbex maculatus*	OL549450	Current study
CSCS-Hym-MC0132		*Asicimbex concavicaputus*	OM066096	Current study
		Asitrichiosoma anthracinum	KT921411	Song et al. (2016a)
CSCS-Hym-MC0035		*Cimbex luteus*	OL549453	Current study
CSCS-Hym-MC0009		*Labriocimbex sinicus*	MH136623	Yan et al. (2019)
CSCS-Hym-MC0034		*Leptocimbex allantiformis*	OL549455	Current study
CSCS-Hym-MC0162		*Leptocimbex clavicornis*	MT478109	Cheng et al. (2021)
CSCS-Hym-MC0168		*Leptocimbex linealis*	OM066094	Current study
CSCS-Hym-MC0166		*Leptocimbex feipu*	OM066093	Current study
CSCS-Hym-MC0167		*Leptocimbex praiaformis*	MT478110	Cheng et al. (2021)
CSCS-Hym-MC0133		*Leptocimbex yanniae*	MT478111	Cheng et al. (2021)
CSCS-Hym-MC0381		*Odontocimbex svenhedini*	OM066098	Current study
CSCS-Hym-MC0161		*Palaeocimbex crataegum*	OM066091	Current study
CSCS-Hym-MC0049		*Praia tianmunica*	MT665975	Cheng et al. (2020)
CSCS-Hym-MC0330		*Pseudoclavellaria amerinae*	OL549456	Current study
CSCS-Hym-MC0165		*Trichiosoma vitellina*	MN853777	Chen et al. (2020)
	Corynidinae	*Corynis lateralis*	KY063728	Doğan & Korkmaz (2017)
CSCS-Hym-MC0176		*Corynis zhengi*	OL549451	Current study

ChinaXiv:202303.00021v1

Table S4. Prior distributions of calibration pionts

Taxon (Crown)	Scheme 1	Scheme 2	Scheme 3
Cimbicidae	60-135	60-135	60-135
Cimbicinae + Abiinae	/	47-135	47-60
Cimbicinae	27-58	37-47	37-47
Abiinae	14-25	>20	>20

chinaXiv:202303.00021v1

Table S5. The mitochondrial genome base composition of Cimbicidae

Speies	Length (bp)	T%	C%	A%	G%	A+T %	G+C %	AT-skew	GC-skew
Abia berezowskii	15,185	37.92	11.16	43.29	7.63	81.21	18.79	0.066	-0.188
Abia jimeii	15,701	38.58	10.32	43.65	7.46	82.22	17.78	0.062	-0.161
Abia niui	15,755	38.44	10.21	44.20	7.15	82.63	17.37	0.070	-0.176
Agenocimbex maculatus	15,442	36.81	13.49	41.67	8.04	78.47	21.53	0.062	-0.253
Asicimbex concavicaputus	15,404	38.46	11.13	42.78	7.63	81.24	18.76	0.053	-0.187
Asitrichiosoma anthracinum	15,391	37.42	11.48	43.35	7.75	80.77	19.23	0.073	-0.194
Cimbex luteus	15,096	38.00	10.91	43.73	7.37	81.72	18.28	0.070	-0.194
Corynis lateralis	14,881	36.71	11.30	43.84	8.15	80.55	19.45	0.089	-0.162
Corynis zhengi	15,444	36.36	11.97	44.08	7.59	80.45	19.55	0.096	-0.224
Labriocimbex sinicus	15,404	37.72	11.11	43.50	7.68	81.21	18.79	0.071	-0.182
Leptocimbex allantiformis	15,085	37.89	10.73	43.62	7.76	81.50	18.50	0.070	-0.161
Leptocimbex clavicornis	15,253	37.38	11.30	43.88	7.43	81.26	18.74	0.080	-0.206
Leptocimbex linealis	15,238	37.61	11.65	43.10	7.64	80.71	19.29	0.068	-0.208
Leptocimbex feipu	15,135	37.43	11.38	43.64	7.55	81.07	18.93	0.077	-0.202
Leptocimbex praiaformis	15,055	37.01	11.64	43.84	7.51	80.85	19.15	0.085	-0.216
Leptocimbex yanniae	15,259	38.25	10.65	43.44	7.67	81.68	18.32	0.064	-0.163
Odontocimbex svenhedini	15,384	37.29	11.65	43.54	7.53	80.82	19.18	0.077	-0.215
Orientabia sinica	15,655	38.36	10.64	43.56	7.44	81.92	18.08	0.064	-0.177
Palaeocimbex crataegum	15,941	38.23	11.05	43.00	7.72	81.23	18.77	0.059	-0.177
Praia tianmunica	15,556	38.22	10.83	43.60	7.35	81.82	18.18	0.066	-0.192
Pseudoclavellaria amerinae	15,181	36.66	11.53	44.36	7.44	81.02	18.98	0.095	-0.216
Trichiosoma vitellina	15,245	37.44	11.05	44.18	7.34	81.61	18.39	0.083	-0.202
Zaraea akebiae	15,841	38.88	9.97	43.80	7.34	82.68	17.32	0.060	-0.152
Zaraea mengmeng	15,798	37.07	11.93	43.21	7.80	80.28	19.72	0.077	-0.209
Zaraea zhui	14,989	37.36	11.06	43.85	7.73	81.21	18.79	0.080	-0.177

chinaXiv:202303.00021v1

Table S6. Genome Assembly and Gene Prediction Statistics for the Cimbicidae

Species	Total data (Gb)	Average depth (x)/Coverage	Total length /size (Mb)	Scaffold number	N50 length (kb)	Longest scaffold (kb)	GC%	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragmented BUSCOs (F)	Missing BUSCOs (M)	Total BUSCO groups searched (T)
Abia berezowskii	7.28	47.30	154.00	14815	21.38	224.84	37.26	5654	32	43	262	5991
Abia jimeii	12.49	73.42	170.17	32388	8.28	364.93	37.40	4860	33	197	901	5991
Abia niui	10.65	63.17	168.61	21877	15.97	300.48	36.99	5367	32	109	483	5991
Agenocimbex maculatus	7.22	46.46	155.32	18756	13.40	362.44	39.90	5386	33	83	489	5991
Asicimbex concavicaputus	12.89	80.20	160.74	9413	31.38	943.16	36.98	5652	36	38	265	5991
Cimbex luteus	2.00	16.77	119.28	148578	1.22	14.81	38.73	1234	5	1397	3355	5991
Corynis zhengi	57.65	349.10	165.14	18988	16.80	312.08	37.40	5660	40	31	260	5991
Labriocimbex sinicus	4.49	25.68	174.98	34297	8.09	134.35	38.81	5338	34	119	500	5991
Leptocimbex allantiformis	1.67	9.17	181.98	102216	3.78	36.58	39.78	3835	14	1068	1074	5991
Leptocimbex clavicornis	11.34	65.57	172.90	24851	11.97	152.44	39.40	5423	43	88	437	5991
Leptocimbex linealis	8.86	48.20	183.81	17047	22.64	269.08	38.96	5686	32	32	241	5991
Leptocimbex feipu	10.10	54.86	184.08	11008	41.35	579.19	38.97	5751	26	21	193	5991
Leptocimbex praiaformis	11.42	64.44	177.27	10181	41.24	516.72	39.32	5786	37	18	150	5991
Leptocimbex yanniae	18.94	104.15	181.84	23988	12.57	146.85	39.13	4732	21	592	646	5991
Odontocimbex svenhedini	15.91	91.00	174.82	12206	30.24	250.63	42.65	5527	24	193	247	5991
Orientabia sinica	16.03	101.69	157.64	27799	9.45	159.92	37.14	5263	30	98	600	5991
Palaeocimbex crataegum	11.17	72.83	153.35	12609	24.70	249.82	37.64	5680	34	31	246	5991
Praia tianmunica	2.51	12.04	208.29	177647	14.23	169.82	38.35	5218	19	390	354	5991

chinaXiv:202303.00021v1

Pseudoclavellaria amerinae	14.06	92.98	151.18	19383	16.21	235.48	40.69	5441	24	243	283	5991
Trichiosoma vitellina	11.93	65.55	181.96	26807	10.93	366.99	38.30	5186	31	145	629	5991
Zaraea akebiae	14.15	83.82	168.85	13380	30.21	345.85	35.09	5430	23	238	300	5991
Zaraea mengmeng	7.60	47.72	159.34	6314	85.05	1027.89	38.04	5796	26	17	152	5991
Zaraea zhui	12.62	79.65	158.42	26938	10.00	285.73	36.15	5258	32	144	557	5991

chinaXiv:202303.00021v1

Table S7. Summary of BUSCO amino acid matrices for phylogenetic analyses

Matrix	Alignment length	Minimum occupancy per locus (%)	Number of loci	Average missing taxa per locus (%)	Number of sites	Missing sites (%)
matrix100	233883	100.00%	602	0.00%	5379309	0.00%
matrix100abs70	176305	100.00%	393	0.00%	4055015	24.61%
matrix90	1645109	91.30%	3381	5.21%	37837507	0.00%
matrix90abs80	740516	91.30%	1044	5.68%	17031868	54.98%

chinaXiv:202303.00021v1